Vol. III Part 3 August 1951

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY

Editorial Board

D. RUSSELL DAVIS
R. C. OLDFIELD J. W. WHITFIELD
G. C. DREW M. D. VERNON
G. C. GRINDLEY O. L. ZANGWILL

CONTENTS

C. B. GIBBS. Transfer of Training and Skill Assumptions in Tracking Tasks - - - - - - - - 99
J. SZAFRAN. Changes with Age and with Exclusion of Vision in Performance at an Aiming Task - - 111
K. DANZIGER. The Operation of an Acquired Drive in Satiated Rats - - - - - - - - - - - 119
V. R. CANE and V. HORN. The Timing of Responses to Spatial Perception Questions - - - - - - 133
N. HARRIS. A Note on the Paired Comparisons Method of Rating - - - - - - - - - - 146
Book Review - - - - - - - - - - - - 148

PUBLISHED FOR
THE EXPERIMENTAL PSYCHOLOGY GROUP
BY
W. HEFFER & SONS LTD. CAMBRIDGE

AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 Massachusetts Avenue, N.W., Washington 5, D.C.

PUBLICATIONS:

AMERICAN PSYCHOLOGIST

Editor: FILLMORE H. SANFORD, *American Psychological Association.* Contains all official papers of the Association and articles concerning psychology as a profession; monthly

Subscription: $7.00 (Foreign $7.50). Single copies, $.75.

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

Editor: J. MCV. HUNT, *Institute of Welfare Research, New York City.* Contains original contributions in the field of abnormal and social psychology, review, and case reports; quarterly.

Subscriptions: $6.00 (Foreign $6.50). Single copies, $1.75.

JOURNAL OF APPLIED PSYCHOLOGY

Editor: DONALD G. PATERSON, *University of Minnesota.* Contains material covering applications of psychology to business, industry, and education; bi-monthly.

Subscription: $6.00 (Foreign $6.50). Single copies, $1.25.

JOURNAL OF COMPARATIVE AND PHYSIOLOGICAL PSYCHOLOGY

Editor: HARRY F. HARLOW, *Department of the Army, Washington, D.C.* Contains original contributions in the field of comparative and physiological psychology; bi-monthly.

Subscription: $7.00 (Foreign $7.50). Single copies $1.25.

JOURNAL OF CONSULTING PSYCHOLOGY

Editor: LAURANCE F. SHAFFER, *Teachers College, Columbia University.* Contains articles in the field of clinical and consulting psychology, counseling and guidance; bi-monthly.

Subscription: $5.00 (Foreign $5.50). Single copies, $1.00.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

Editor: ARTHUR W. MELTON, *U.S. Air Force, San Antonio, Texas.* Contains original contributions of an experimental character; monthly, two volumes per year.

Subscription: $14.00 for 2 vols. (Foreign $14.50). Single copies, $1.25.

PSYCHOLOGICAL ABSTRACTS

Editor: C. M. LOUTTIT, *University of Illinois.* Contains non-critical abstracts of the world's literature in psychology and related subjects; monthly.

Subscription: $7.00 (Foreign $7.50). Single copies, $.75.

PSYCHOLOGICAL BULLETIN

Editor: LYLE H. LANIER, *University of Illinois.* Contains critical reviews of psychological literature, methodological articles, book reviews, and discussions of controversial issues; bi-monthly.

Subscription: $7.00 (Foreign $7.50). Single copies, $1.25.

PSYCHOLOGICAL MONOGRAPHS: GENERAL AND APPLIED

Editor: HERBERT S. CONRAD, *U.S. Office of Education, Washington, D.C.* Contains longer researches and laboratory studies which appear as units; published at irregular intervals, about ten numbers per year.

Subscription: $6.00 per volume (Foreign $6.50). Single copies, price varies according to size.

PSYCHOLOGICAL REVIEW

Editor: CARROLL C. PRATT, *Princeton University.* Contains original contributions of a theoretical nature; bi-monthly.

Subscription: $5.50 (Foreign $6.00). Single copies, $1.00.

THE QUARTERLY JOURNAL OF
EXPERIMENTAL PSYCHOLOGY

Vol. III 1951 Part 3

TRANSFER OF TRAINING AND SKILL ASSUMPTIONS IN TRACKING TASKS

BY

C. B. GIBBS

(*From the Medical Research Council Applied Psychology Unit, Cambridge*)

Matched groups of subjects were used to test the learning and transfer effects that follow changes in the display, the muscular reactions and the directional relationship between stimulus and response in a tracking task. Two arrangements were compared in the relationship studies: one arrangement of the stimuli and reactions was similar, and the other was opposed to that used in many every-day skills. The familiar arrangement was easier to learn. There was high positive transfer from the unfamiliar to the familiar, and little transfer from the familiar to the unfamiliar.

The physical dimensions of the display were varied to give two tasks with different stimuli. The initial learning times were equal for both tasks, and the transfer between them was high, positive, and equal. Two further tasks varied in the extent, speed and force of the required muscular movements. One task proved more difficult to learn initially, and there was greater transfer from the difficult to the easy task than from the easy to the difficult. A further experiment tested the effects of changing the difficulty of a tracking course, and it was found that learning was more rapid on the more difficult course. A difference in difficulty between two tasks, therefore, determined both the amount of transfer between them and the rate of learning the tasks.

New measures were developed to test the transfer between tasks of unequal content, and the effect of such inequalities upon the rate of learning. The findings are discussed, as are their possible implications for transfer measurement and their bearing upon existing theories of transfer.

I

Introduction

The theory of identical elements appears to dominate the analysis of the problems of transfer, together with the view that transfer depends, in some way, upon the similarity of the stimuli and responses in the first and second task, as measured along some physical dimension. This theory is no longer adequate by itself, but it summarizes some major trends in the experimental data. It has been demonstrated, for example, that in some tasks the physical characteristics of the stimuli and muscular reactions may be altered over a certain range, but these differences do not alter the nature of the tasks so far as learning and transfer are concerned; and the subject can readily transfer from any task within this range to any other task that also lies within this range. Bartlett (1947) has pointed to similar zones of equivalence in performance tests, and has developed the concept of *limits of ability* to define these zones. Here the physical variables of a task may be altered over a certain range without affecting the standard of performance. It is possible that there is a close relationship between the zones of equal performance and those of equal transfer, and the problems of transfer measurement and prediction would be greatly simplified if such a relationship could be demonstrated.

Such a connection may well exist in certain tasks which vary only in the physical dimensions of the stimuli and reactions, but these are not the only elements that make up the total content of a task, and it is here that the second concept is introduced. The stimuli and reactions may be related in a way that is expected or unexpected in relation to the experience of the subject before the experiment, and this factor may dominate the initial learning and the transfer between two tasks. This is a most important aspect of learning and transfer research which has so far been greatly neglected, apart from the preliminary work of Vince (1944), Warwick (1947) and Mitchell (1947) which has been summarized by Mitchell and Vince (1950). From previous experience there are built up a large number of *skill assumptions*; the subject may be quite unaware of the existence of them although they are used many times in the day. He will expect that, if he moves a control lever to the right, then some element in the visual display will respond by moving to the right rather than to the left. This simple assumption is clearly based on long experience dating from infancy. But, starting much later in life, it is possible to build up extremely extensive and complex skill assumptions as, for example, any car driver finds he has adopted, when he changes from a country with left-hand to one with right-hand rule of the road. Then, again, much shorter term skill assumptions are found when a subject begins to expect from his recent immediate experience of a changing display, that the signals will subsequently follow the initial patterns they presented in space and time (Poulton, 1950). A similar static instance occurs when one expects to find stationary objects where they were last seen.

The real importance of skill assumptions is found in transfer experiments when one of the tasks is in conflict with everyday experience. The expected task will undoubtedly be easier to learn, whether it is given before or after the unexpected task. In the extreme case, for example, even if there were no learning during the initial unexpected task of a transfer experiment, there might well be apparently complete "transfer" towards a task organized in a way that everyday experience had taught long before the subject had entered the laboratory.

Two procedures were used in these investigations:

1. The handle-winding task.
2. The steering task.

II

Handle-winding Experiments

Procedure.

Apparatus. The task given was that of compensatory tracking, and control was by handwheel. An electric motor tended to rotate a pointer at a constant rate away from a datum position. The pointer tip was visible to the subject, and his handwheel control was connected to the pointer through a mechanical device, that added or subtracted the handwheel drive and that of the electrical motor.

The result was that when the subject turned the handwheel continuously and smoothly at rates of 18 turns per minute (i.e. at the same rate as the electric motor, but in the opposite direction) the two drives cancelled out, and the pointer tip remained stationary against the index mark.

This was the operator's task, and any pointer movements were therefore errors. These degrees of rotational error were summed on a further dial and recorded every minute.

Practice periods and criterion. The subjects practised winding continuously for five-minute periods with a three-minute rest period between each practice. Only three such periods of practice were given in any one forenoon or afternoon. The criterion of complete learning adopted and used for all groups was that errors should not exceed 5° over the

whole five-minute period. Scores at this level were then examined in detail to determine learning times to the nearest minute, since a score could be less than 5° over a five-minute period, but the score for any one minute might exceed 1°. This could occur at any time between the first and fifth minutes, and the learning scores were summed to the minute when such errors last occurred. In effect, the 5° criterion determined learning times to the nearest five minutes, and the 1° standard defined the nearest minute. After attaining this level of ability on the first task, all subjects and all groups continued practice on their initial task, to a total of 40 minutes' practice, to equate practice for all groups before transfer.

They then transferred to their second task, and the time to reach the criteria on this task was recorded, again to the nearest minute.

Variations of the apparatus. The speed of the electric motor remained constant in revolutions per minute throughout all the tests, but nevertheless a change in the apparatus that substituted a different *length of pointer*, resulted in an entirely different velocity of the pointer tip, for any given error in the winding rate of the handwheel. The stimuli were therefore altered. The corrective movements of the handwheel, that were associated with particular errors were also altered, i.e. the ratio of hand to pointer movements, or gear ratio, was also varied by this change.

Changing the *handwheel size* altered the basic speed and distance moved by the hand in maintaining a constant rate of winding. The ratio of hand to pointer movements (gear ratio) was also altered. The hand moved through different distances and at different speeds, to bring about any given correction at the pointer. A further change in the physical conditions also resulted, because a different basic force was needed to rotate the handwheel, when the handwheel diameter was changed.

The *direction of winding* could also be easily reversed. One arrangement needed a constant rate of *clockwise winding*, but, by changing the direction of rotation of the electrical motor, *counter-clockwise* winding was required from the subject.

The *sense of the directional relationship* between handwheel and pointer could also be varied. Clockwise handle-winding could result in clockwise pointer movements, or, by using a reversing gear in the drive between handwheel and pointer, in counter-clockwise movements of the pointer. (Fourteen subjects were used in a preliminary experiment to determine the preferred or expected relationship. Their responses showed that 12 subjects expected that a clockwise handwheel movement would result in clockwise pointer movement.)

Five out of all possible arrangements of these variables were selected for study:

1. *Standard arrangement.*—The standard arrangement of the apparatus; this was:
   (i) Small handwheel of 3-inch diameter, requiring a basis force of 0·75 lb.;
   (ii) Small pointer of 3-inch radius;
   (iii) Clockwise direction for handle-winding;
   (iv) Expected directional relationship; clockwise rotation of hand giving clockwise pointer movement.

2. *Unexpected arrangement.*—This changed only one factor in the standard arrangement—the directional relationship; clockwise movement of the hand gave a counter-clockwise movement of the pointer.

3. *Large pointer arrangement.*—Also differed from the standard arrangement in only *one* respect, the large 12-inch pointer was used instead of the small 3-inch pointer. This quadrupled the speed error of the pointer tip that resulted from any given error in winding rate and altered the gear ratio by 4 : 1.

4. *Counter-clockwise arrangement.*—Again this altered one factor in the standard arrangement; counter-clockwise turning was required instead of clockwise.

5. *Large wheel arrangement.*—Was also different in one respect from the standard arrangement; the large 6-inch diameter wheel was used instead of the small 3-inch wheel. The extent and speed of hand movement was doubled, and the basic winding force was halved. The gear ratio was altered in the ratio of 2 : 1.

The matching and allocation of the groups.

Altogether 70 experimental subjects were tested, all naval ratings aged between 19–25 years. By giving everyone a preliminary practice spell of one minute on the standard arrangement (1) of the equipment it was possible to divide the 70 subjects into seven

groups of ten men each (Groups A, B, C, D, E, F and G) and to ensure that the average initial ability of any one of these groups was the same as that of any other.

Two of these groups were used to compare the effects of any one factor; one group learned the standard, followed by the modified arrangement; with a second group this order was reversed.

III

Results of Handle-winding Experiments

The average learning times for each group of subjects were used to compare the effects of the different experimental arrangements. The data have been summarized in Table I.

TABLE I

Average Time (in minutes) to Learn the First and Second Tasks

Nature of problem	*Group*	*Initially learned arrangement*	*Initial learning time*	*Transfer to arrangement*	*Transfer learning time*	*Total time to learn both arrangements*
Change of directional relationship	A	Expected	20·5	Unexpected	25·7	46·2
	B	Unexpected	30·0	Expected	1·9	31·9
Change of pointer size	C	Small	18·3	Large	2·7	21·0
	D	Large	20·0	Small	2·6	22·6
Change of direction of winding	E	Clockwise	19·0	Counter-clockwise	2·3	21·3
Change of handwheel size	F	Small	19·6	Large	1·0	20·6
	G	Large	11·1	Small	3·7	14·8

Change of directional relationship.

The results in Group A and Group B showed that the change from the unexpected to the expected arrangement was much more readily undertaken than was the change from the expected to the unexpected. The preliminary practice of Group A on the expected arrangement did not reduce the time required to learn the unexpected arrangement (25·7 minutes) by a significant amount compared with the average time for Goup B (30·0 minutes), who learned the unexpected arrangement first. ($t = 1{\cdot}05$—not significant. A value for t of 2·101 was required in all these experiments to show a significant difference at the $P = 0{\cdot}05$ level.)

On the other hand, Group B learned the expected arrangement very rapidly after their previous practice on the unexpected arrangement. This time (1·9 minutes) was compared with that of Group A (20·5 minutes), who learned the expected arrangement first ($t = 7.52$, indicating a high degree of significance). Learning the unexpected arrangement first therefore saved 18·6 minutes in learning the expected arrangement.

The total time to learn both the unexpected and the expected arrangements was markedly different in the two orders; practice on the unexpected relationship first (Group B) meant that only 32 minutes were required to learn both tasks, whereas 46 minutes were needed by Group A ($t = 2{\cdot}86$).

The initial learning times showed that the unexpected arrangement was definitely more difficult since Group B took 30 minutes, instead of the 20 minutes required by Group A for the expected arrangement ($t = 2{\cdot}54$).

Change of pointer size.

The results in Groups C and D showed that altering the gear ratio, and hence the visual stimulus in magnitude or speed, gave a high positive transfer. Transfer was substantially the same whether it was in the direction large pointer to small pointer, or in the reverse order, and total times to learn the two tasks were not affected by the sequence in which they were learned, since these times were 21·0 and 22·6 minutes respectively. In both groups the amount of skill transferred was large, positive and highly significant. A comparison of the time required by Group C to learn after transfer (2·7 minutes) with the time taken by Group D (20 minutes), who learned the same form of the task as an initial task, gave $t = 6{\cdot}96$ for transfer from small to large stimulus. Similarly, a t value of 5·56 was found for transfer from large pointer to small, the average times here being 2·6 minutes to learn after transfer, and 18·3 minutes with no preliminary training. There was no significant difference in the initial learning times of the two tasks. It was evident therefore that changes in visual stimulus and gear ratio did not affect learning times, rate of learning, or amount of skill transferred.

Change of direction of winding.

The results in Group E showed good transfer between two arrangements where response was altered from clockwise to counter-clockwise turning; only 2·3 minutes were required to learn the second task after 19 minutes had been spent in learning the first task.

Change of handwheel size.

The initial learning times of Groups F and G showed that there was a significant difference in the times required for the initial learning of the two tasks differing in the size of handwheel ($t = 2{\cdot}85$). The small handwheel took 19·6 minutes to learn, and the large took only 11·1 minutes. This difference in difficulty between the two tasks was not associated with any reliable advantage in presenting one task before the other, in order to reduce the total time to learn the two tasks; the difference between the two sequences of presentation gave a t of 1·67, which was not significant at the $P = 0{\cdot}05$ level.

On the other hand, there was greater transfer from the small handwheel to the large than in the opposite order of learning, since the large handwheel only required an average of one minute to learn, after the preliminary practice on the small handwheel; whereas 3·7 minutes were required to learn the small handwheel when preliminary practice had been on the large. This was a significant difference, t being 3·7. Learning either arrangement first resulted in high positive saving on the other as a second task, and the benefits of the preliminary training were highly significant in both cases ($t = 6{\cdot}3$ and $t = 5{\cdot}0$).

IV

Steering Task

Procedure.

The task of serial pursuit was set in this experiment, and the subjects tracked along a moving line of electrical contact studs with a pointer carrying a single ball contact. This pointer was moved by a handwheel, and the task was like steering along a winding road. One course was simple to follow, with few curves, whereas the other was difficult, with frequent and sharp curves. It was relatively easy to anticipate correctly in one task and also easier to memorize the course. In both courses the subjects could only see two contact

studs in advance of the pointer, and the rest of the course was hidden. Visual anticipation time was constant in both courses, but one course was more like that expected by most people in continuous tasks, i.e. that the signals remain relatively constant, and there is a good deal of repetition (C. Poulton, 1950). One group had all their initial practice of 21 runs on the difficult course and were then transferred to the easy course. This Group (H) was made up of 10 business women of ages 32–52. The other Group (I) learned the same tasks but in the opposite sequence. They were 10 naval ratings of ages 19–25 years, and they were considerably better than the women at a short preliminary test of the same kind of skill. It was intended to see whether a good system of training could make a poor group more accurate than a good group provided with only an indifferent training.

The method of scoring was that contacts, or "hits," were only made when the subject kept the pointer exactly in line with the moving line of contact studs. These hits were added by counter to give the score for any one run. The highest possible score on either course was 194 hits. Two points in each learning curve were taken to compare the groups on the steering tests. These were the group score on their 21st and final run on the first task, and their first score on the second task. These scores were directly compared, and they were also added to compare the overall ability of the two groups at the same stages in their training.

V

Results of Steering Experiment

The group average scores are set out in Table II.

TABLE 2

Results of Steering Experiments

Group	*Final score on first task*	*Initial score on second task*	*Sum*
H: Business women—difficult task first easy task second	165	182	347
I: Naval ratings—easy task first difficult task second	181	144	325

In spite of their low initial ability the training of Group H on the difficult task resulted in a more rapid extension of their abilities, and at the end of their training they were superior to Group I, on both the task that they had practised and the final task that was new to them. Group H averaged 347 hits for the difficult plus the easy run, whereas Group I averaged only 325 hits; this difference was highly significant, t being 3·5. It is of particular interest that the first score of Group H on transfer to the easy task was the same as the final score of Group I on the easy task, although the latter had practised this task *directly* for 21 runs, with all the advantages of memorizing that this implies. Indirect was therefore equivalent to direct practice. This is an interesting finding, and when the innate inferiority of Group H and the greater memorizing advantages of the younger Group I are taken into account, it seems likely that indirect practice was possibly superior to direct, although this has not yet been proved. It follows, however, that a carefully adjusted training can greatly accelerate learning rates, even in a poor group, and show a very real extension of general ability, as compared with a superior group who were not exercised at full capacity.

Transfer effects were high and positive in both cases. But initial training on the easy task gave much less transfer to the difficult task, and the inferiority of direct to indirect practice in this case was highly significant ($t = 4{\cdot}3$). Group H on the difficult

task scored 165 hits after 21 runs of direct practice, and Group I scored only 144 hits in their initial run on the difficult task, after their preliminary training on the easy task.

The apparent difference in difficulty was confirmed, since Group H increased their score from 165 to 182 on transferring from the difficult to easy track ($t = 6{\cdot}7$), and it has been shown that practice on the difficult task was equal to practice on the easy task. Also Group I scores decreased from 181 hits on the easy task to only 144 hits on the difficult task ($t = 8{\cdot}5$).

VI

Discussion

Test design for handle-winding experiments.

A new test design is needed to study the effects upon transfer of skill assumptions or unequal task content, since the standard designs use equal tasks, or pool the data when unequal tasks are used, and are not suited to the experimental purpose. The conventional transfer formulæ cannot be used to show the transfer interactions between unequal tasks, because the percentages are determined by two factors: the amount of skill carried over from the first to the second task, and the content of the second task. If two tasks are unequal, and are presented in opposite order, then the calculated transfer percentages must also tend to be unequal; but it cannot be assumed from this that the transfer of skill is different in the two orders of presentation since two effects are confounded in the percentages formulæ. Thus the measures of the average learning time for each task, and the total learning time for each pair of tasks, have been used in the handle-winding experiments. The former measure can be used to show the time saved in learning each second task because of the preliminary training on the first. The levels of task difficulty, and the amounts of skill transferred, can therefore be related to each other. The latter measure shows the relationship of task difficulty and the order of presentation, to the total learning time for both tasks.

The total learning time is an essential measure. It is evident that time will be lost in learning the more difficult task first, but there may be a compensating saving in learning the easy task after transfer. There may be an equal, but opposite effect in the reverse order of learning; so that the total learning time is the same and unaffected by the order. This is defined as simple additive compensation.

Other tasks may show a difference in total learning times, depending upon the learning orders, and such tasks must be identified and treated separately.

Test design for steering experiments.

The steering test did not proceed to complete learning of either task. The test was designed to compare the effects of task content upon the *rate of learning* two tasks based upon a common skill, rather than the absolute amount of transfer that might have been found if learning had been complete.

Simple relationship between task content and amount of skill transferred.

The handwheel-size and pointer-size experiments (Table I) show that there was more transfer from the difficult to the easy task than from the easy to the difficult, and equal transfer between the tasks when they were of equal difficulty. The unequal transfer of the handwheel-size experiment cannot be explained by theories which state that the amount of transfer is related to the degree of similarity of the stimuli and responses along some physical dimension. It is clear that the separation of the

physical factors is the same no matter in which order two tasks are presented, and transfer would be equal if it depended only on physical separation. A new approach has to be made to the problems of generalization and discrimination in transfer.

The stimuli and reponses were changed substantially in the experiments on pointer and handwheel size, but there was high positive transfer between all such variations. It follows that when one arrangement of the physical variables was learned, this learning extended over a wide range of these variables. The pointer-size experiment gives the simplest case, with equal learning times, and very high transfer in both directions between the two arrangements, in spite of considerable alterations in the stimuli. The two arrangements must indeed have been virtually the same task, from a psychological point of view, in spite of large physical differences.

The concept of different levels of required ability.

This indicates that a concept of *levels of required ability* may be important in transfer. Bartlett (1947) has outlined the concept of limits of ability. He pointed out that wide variations can be made in physical factors without any corresponding change in human performance, provided these changes lie within a certain range on the physical scale of the environmental factors in question. "The fundamental features of a performance will remain stable over a certain range of its conditions. Outside this range they will change often in a dramatic and radical manner." A major purpose of performance and learning tests is to determine the range over which a physical condition can be varied without affecting performance or learning scores. When the physical variables are altered, the tolerance limits can be defined, beyond which the task will show an increase in difficulty. Tasks within an optimal zone are psychologically equivalent in difficulty and in the level of ability they require; a common measure of such zones is that a given proportional change in the stimuli or responses is psychologically equivalent, e.g. the Weber or Fullerton Cattell Square Root Law applies. The results of the present transfer experiments suggest that tasks within such limits may not only be equal in performance and in required ability, but also in transfer, i.e. there is high and equal transfer between any two tasks within these limits. This finding deserves extensive study to discover how general it is.

Tasks outside the optimal zone require superior ability to reach a given standard of performance than tasks within the zone. It is clear that the level of ability gained in practising a task must be limited by the content of that task; thus the ability to deal with difficult situations cannot be acquired in practising easy ones. The present results could be explained therefore by a hierarchy of levels of required ability, with the higher levels including the lower. Practising a task to total possible skill may develop a level of ability that is directly related to the difficulty of the task, and there would then be high positive transfer from difficult to easy tasks and less transfer in the opposite direction. The higher levels of ability could then deal with a larger range of physical variables in the stimuli and responses than could the lower.

Relationship between measures of performance and of transfer.

The connection between learning and performance tests on the one hand and transfer tests on the other has been emphasized, since if such a connection exists then there is little point in carrying out laborious transfer tests on certain classes of physical variables, e.g. when it is known that a range of such variables conform to Weber's Law. There would also be important implications for applied work if the data from some types of performance tests could be used in the practical problems of transfer in the field.

It is clear that, when there is simple additive compensation in transfer between any two tasks, there is a direct relationship between such transfer tests and performance and learning tests that use the same experimental material. This must follow because the first part of a transfer test (the initial learning or performance test on two tasks) is in itself a performance and learning test. If there is simple additive compensation, the difference between the scores on the second tasks is equal to the difference between the scores on the first tasks, and a simple relationship must therefore exist between such transfer results and performance and learning tests. It may be important therefore to isolate the class of effects that lead to simple additive compensation in transfer, and to compare the present transfer results with independent learning and performance tests that have used the same experimental variables.

The studies of changed pointer size show equal times to learn the large or small pointer arrangement, and equal transfer from small to large, and from large to small. The change of pointer size altered the speed of the pointer tip and the gear ratio. Hick (1950) has shown that in judging stimulus speed there is an optimal range of speeds that conforms to Weber's Law. Vince (1946) has shown that tracking with a continuously changing gear ratio gives a performance score that is equal to a constant ratio.

There is also the evidence from the changes in handwheel size. The large handwheel was easier to learn as an initial task, but there was less transfer from large to small. The effects were approximately compensatory, however, since there was not a significant difference in the total time to learn the two tasks in the different orders of presentation. The basic physical force was lower, and the tangential speed of the hand was higher when the large handwheel was used, and these were nearer to the optimal conditions that have been defined in independent performance tests, *cf.* Hick (1945) and Helson (1949). These two different tasks had different levels of difficulty, as shown in the initial test, and transfer was found to be approximately proportional to the difference in difficulty.

It is evident that the results of the above transfer tests could have been predicted, with fair accuracy, from the independent performance tests. These tests altered only the physical dimensions of the basic stimulus and response, and the tasks required the same kind of skill. It is important to note that learning approximated to total possible skill in all tasks before the groups were transferred. These are essential conditions for the hypotheses that follow.

The results of the pointer size studies suggest the hypothesis (i) that *transfer between two tasks, both of which lie between the same tolerance limits, as defined by performance tests, can be* (a) *positive,* (b) *large, and* (c) *equal and unaffected by the order of task presentation.*

The handwheel results suggest the hypothesis (ii) that transfer between two tasks, one of which lies within the tolerance limits, and one of which lies outside this optimal zone can be (a) *positive,* (a) *large,* (c) *unequal and affected by the order of task presentation; this implies greater transfer when the first task lies outside the optimal zone, and the second lies within it, than when the first task comes within and the second outside.*

The changes in directional relationship between handwheel and display.

The change in directional relationship gave results that accord with hypothesis (ii), since the transfer from the difficult and unexpected relationship to the expected was very much greater than the transfer from the easy to the difficult task. This result also supports the concept of a hierarchy of levels of required ability rather than that of physical similarities, but there are several reasons why this study must be treated separately.

It is important to note that the basic stimuli and responses were the same in both the unexpected and expected versions of this task, i.e. they were common elements in both tasks. Nevertheless, the transfer from the expected to the unexpected organization of these elements was very small and had no significant effect in reducing the time to learn the latter task. This suggests that the elements of stimuli and responses may have comparatively little effect in transfer, whereas their relationship in space and time may be all important. Indeed, Bartlett (1947) has previously made the same point that stimuli and muscular reactions by themselves are not meaningful elements in the analysis of given tasks, whereas the organization of these in space and time may well be all important. This view is supported by comparing all the transfer effects of the handle-winding experiment in Table I. The change in directional relationship had a far greater effect on transfer than any of the isolated changes in stimuli or muscular reaction.

The effect of changed time-space relationships upon transfer.

It was found in the handle-size experiment that the low level of ability that was gained in practising an easy task made some contribution to the higher level of ability that was required in a more difficult task, and this contribution was roughly proportional to the difference in the difficulty of the tasks. This was not the case in the directional-relationship studies, where learning the easy task first did not result in any significant saving in learning the difficult task after transfer, whereas learning the difficult task first gave very high transfer to the easy task. There is a simple explanation of this finding, since if the essential organization of the expected arrangement had already been learned and skill assumptions built up before the experiments began, then there would be the appearance of positive transfer to it as a second task, irrespective of the relevance of the learning that took place during the course of the experiment.

Conversely, if the unexpected organization was not learned in everyday experience and was not present in the expected arrangement (that was learned as a first task by Group A during the experiment), then no knowledge of this unfamiliar arrangement could have been carried over from the expected to the unexpected arrangement or from previous experience. The learning of the unexpected arrangement had to start virtually from scratch, in spite of the preliminary training on the expected arrangement.

The skill assumption that was present in the expected directional relationship, and carried over from everyday experience, was that the subjects expected that any given reaction would move some element in the display in the same sense as the response. This assumption is clearly developed in the co-ordinations that are used in writing, pointing at a moving object, and in other everyday skills. Many other skill assumptions are to be found in serial motor skills, and several examples have been given in the Introduction. Their effect is apparently to form a preference for certain forms of organization of the stimuli and responses in space and time. It seems unlikely that simple additive compensation will be found when skill assumptions prove to be correct in one task, but not in the other. The effect is that transfer scores cannot be easily predicted from performance data in these cases, although it is conceivable that such prediction can be made when the physical dimensions of the stimuli and responses are the only elements that are altered.

It may be generally true, however, that, if new stimuli and responses are learned in a familiar relationship, this learning will make little contribution toward learning a subsequent task, where the same stimuli and responses are organized in direct conflict with the familiar relationship. There may indeed be negative transfer in some cases, since the initial learning might reinforce skill assumptions that have to be discarded in the second task.

The differences in total learning times.

The results show that in all cases it was better to teach the more difficult task first to obtain the greatest *amount* of transfer between the tasks.

This is not always the case if the total learning time or *rate of learning* both tasks is taken as the criterion. In the handwheel-size experiments there was not a shorter total learning time when the difficult, small handwheel was presented first, in spite of the greater transfer from the difficult to easy task.

The directional relationship experiments with the handle-winding task and the steering experiment showed nevertheless a higher overall rate of learning when the more difficult task was presented first. This is shown by the superiority of total scores with this order. It is clear that the rate of learning is affected by the content of the initial task, but an increase in task difficulty does not automatically increase the rate of learning. It is more probable that there is an optimal level of task difficulty, and an optimal rate of presenting new items to replace those already learned, for every kind of learning material.

The transfer bias between unequal tasks.

The differential effect of unequal task content upon transfer has been known for a considerable time; hence the directions in the early transfer tests that equal tasks should be used, or that data from unequal tasks should be pooled. It appears that Cook (1937) first realized that the effects of task difficulty merited separate investigation, but it is only recently that interest has been revived in the problem, and this has led to a number of simultaneous and independent investigations—e.g. Szafran and Welford (1950), Gagne *et al.* (1950), Gibbs and Clutton-Baker (1950) and several other British and American psychologists who have not yet published their findings. There is a considerable amount of evidence from these sources, and it is generally found that there is more transfer from the difficult to the easy version of a task than in the opposite order. It is unfortunate that these independent findings cannot yet be directly compared because the effect can arise from different causes. The present test design appears to give a method of analysing the general effect into its separate causes, i.e. by avoiding the confusion arising from the use of transfer percentages and by distinguishing between measures of time saved and of total learning time.

VII

Conclusions

1. A design that estimates the transfer between unequal tasks should use measures that distinguish between the amounts of transfer and the rates of learning, e.g. between time saved and total learning time. Transfer should not be expressed in percentages or proportions.

2. It appears that the amount of transfer between two equally difficult tasks may be equal, whereas the transfer between two unequal tasks may be unequal. There may be greater transfer from the difficult to easy task than from the easy to difficult, if the same kind of ability is required in both tasks and learning is carried on until total possible skill is closely approached in both tasks.

3. The effect may be explained by the concept of a hierarchy of levels of required ability; it cannot be explained by the theory that transfer depends only upon the separation of stimuli or reactions along some physical dimension. The detailed causes of the effect may lie in the differences in experimental learning, or in the previous experience of the subjects.

4. A change in the space/time relationship between a stimulus and response may have a greater effect on learning and transfer than a change in the physical dimensions of the stimulus or required response. Learning new stimuli and responses in a familiar relationship may make no contribution to the subsequent learning of these same stimuli and responses in an unfamiliar relationship.

5. It may be generally true that if tasks are of unequal difficulty because of a difference in the physical values of the stimulus or response, then the differences in transfer may have a simple arithmetical relationship to the differences in task difficulty. In these cases, the transfer effects may be predicted, at least approximately, from the data of performance and learning tests that have used the same experimental material.

6. But this simple relationship between task difficulty and the amount of skill transferred is not likely to be found if previous experience aids performance on one task and hinders it on the other.

7. The level of task difficulty affects the rate of learning (i.e. total learning time), but it cannot be held that an increase in task difficulty will inevitably result in an increased rate of learning. It seems more likely that there is an optimum range of task difficulty that ensures the most rapid learning.

My thanks are due to Lt.-Commander Clutton-Baker, who carried out many of the experiments. Dr. N. H. Mackworth made an invaluable contribution by clarifying an initial draft, and other members of the Laboratory made useful suggestions.

References

Bartlett, F. C. (1947). Miscellanea Psychologica. Louvain; pp. 440-452.

Cook, T. W. (1937). Amount of material and difficulty of problem solving. *J. exp. Psychol.*, **20**, 288–296.

Gagne, R. M. and Baker, K. E. (1950). Transfer of discrimination training. *J. exp. Psychol.*, **40**, 314–328.

Gibbs, C. B. and Clutton-Baker (1950). Factors in the initial learning and transfer of serial motor skills and their application to training devices. R.N.P.R.C. 50/572.

Helson, H. (1949). Design of equipment and optimal human operation. *Amer. J. Psychol.*, **62**, 473–496.

Hick, W. E. (1945). The precision of incremental muscular forces with special reference to manual control design. Med. Res. Council A.P.U. Report No. 23.

Hick, W. E. (1950). The threshold for sudden changes in velocity of a seen object. *Quart. J. exp. Psychol.*, **2**, 33–41.

Mitchell, M. J. H. (1947). Direction of movement of machine controls. Ministry of Supply, S.M. 10018 (S).

Mitchell, M. J. H. and Vince, M. A. (1950). The direction of movement of machine controls. *Quvrt. J. exp. Psychol.*, **3**, 24–35.

Poulton, E. C. (1950). Perceptual anticipation in tracking. Med. Res. Council A.P.U. Report No. 118/50.

Szafran, J., and Welford, A. T. (1950). On the relation between transfer and difficulty of initial task. *Quart. J. exp. Psychol.*, **2**, 88–94.

Warrick, M. J. (1947). Direction of movement in the use of control knobs to position visual indicators. U.S.A.A.F., MR & SEAA–694–4C.

Vince, M. A. (1944). Direction of movement of machine controls. F.P.R.C., Report No. 637.

Vince, M. A. (1946). The psychological effect of a non-linear relation between control and display. M.O.S. S.R.I. Report No. 2.

(*Received 1st December*, 1950.)

CHANGES WITH AGE AND WITH EXCLUSION OF VISION IN PERFORMANCE AT AN AIMING TASK

BY

J. SZAFRAN

(From the Nuffield Research Unit into Problems of Ageing, Psychological Laboratory, Cambridge)

Forty industrial workers between 20 and 60 years of age took part in the experiment. Each sat facing a display in an enclosure. His task was to locate targets with a pointer, under two conditons: when he was able to make whatever use he chose of direct vision, and wearing a pair of red goggles through which only the diplay could be seen.

On the average, those over 40 years of age took more time than younger subjects over those portions of the task which were concerned with the initiation of a fresh action. When unable to make use of vision the older had far greater difficulty in locating the targets and attempted to supplement tactile and kinæsthetic cues by making postural adjustments, turning their heads and often their bodies in the direction of the target at which they were aiming. These findings are discussed in relation to an earlier observation of the greater tendency for older people to look at what they are doing.

I

Introduction

Two previous studies by the writer (reported by Welford, 1951) suggested that younger subjects tended to carry out a part of a simple perceptuo-motor task without the aid of visual cues more often than older subjects. Although this occurred during a relatively unimportant phase of an action, there were reasons to believe that as a result the nature of the whole performance was affected. It has further been demonstrated (Welford, 1951) that in studies of ageing measures of overall performance tend to obscure important differences in the constituent reactions of a skill.

Accordingly, the attempt has been made to consider further the importance of visual cues for older people and to study the constituents within a relatively simple skill. Starting with a hypothesis that younger people are more likely than older to be able to carry out accurate limb movements without looking where they are reaching, an experiment was designed to investigate the efficiency with which people of different ages can locate by hand various positions near them in space, with and without the aid of vision.

II

Method

The subject sat in the centre of a wooden framework and faced a display of 19 small electric lights. He was surrounded by a corresponding number of 14-inch square targets, each carrying a round 1-inch diameter metal plate as bull's eye. These targets were arranged in various directions, but at a constant distance, within easy reach, of about 22 inches as measured from the midpoint between his shoulders. Considered vertically, six targets were at the subject's shoulder level, seven 45° above and six 45° below. Considered horizontally, one target was directly in front of the subject and 45° above his shoulder level, three targets 45° to the right side and three 45° to the left, three targets 90° to the right and three 90° to the left, three targets 135° to the right and three 135° to the left. The space directly in front of the subject at 45° below his shoulder level, and also all the space 180° behind him were not used. The space directly in front of the subject and at the level of his shoulders carried the panel of 12 v.6 w. bulbs which were arranged in such a way as to represent a plan of the layout of targets. Half-way between the

subject's seat and the display was a small pillar carrying a 4-inch square metal plate mounted so as to be at the level of the subject's knees.

The subject was instructed to hold a metal stylus ¼ inch in diameter and 6 inches long in a vertical position on the plate in front of him and upon noticing a light on the display being switched off to move the stylus to the centre of the corresponding target, contact with which caused the absent light to reappear on the display. When this was achieved, he was required to return to the starting point.

The centres of the targets, the bulbs which served as target indicators on the display, the plate in front of the subject and the stylus held by the subject were wired to pens recording on a band of paper moving at a constant speed over a drum. These pens registered the following information about each complete operation:—

Time for the initiation of response. The time elapsing between the presentation of the signal and the initiation of the response to it (i.e. lifting the stylus from the plate).

Outward-transit time. The time taken to move the stylus from the plate to make contact with the target indicated.

Search time. The time required to locate the centre after contact had been made with the target area.

Bull's eye time. The time spent in contact with the bull's eye.

Return time. The time taken to bring the stylus back to the starting point.

The apparatus was essentially like that employed by Fitts (1947). However, whereas in Fitts's arrangement the subject's performance was measured in terms of accuracy of his initial contact with each target irrespective of the time taken, the apparatus used in the present experiment measured performance in terms of time taken to achieve a predetermined level of accuracy.

It was explained to the subject that he could carry out the task in any manner he liked, but that he was not allowed to touch any part of the apparatus with his hand. No mention of speed or any other criterion of efficiency was made.

After a demonstration trial every subject carried out the task under two conditions, Condition A always coming before Condition B.

Condition A: the subject was able to make whatever use he chose of direct vision.

Condition B: the subject wore a special pair of red goggles through which he could see the filaments of the bulbs forming the display, but nothing else. Subjects were not told that they would have to face this condition until after they had completed their performance under Condition A.

Every subject was given two series of ten targets in each condition, first holding the stylus in the dominant hand and then holding it in the other hand. For a right hand performance the target directly above the display and the nine targets of the right bank were given. In a left hand performance, the same target above the display and the nine targets of the left bank were employed. The order of presentation of targets was randomized in four Latin squares, allocating a different arrangement to every series in every condition. All the age groups received the same overall treatment.

It was originally intended to compare Condition A with Condition B, which would have necessitated half the subjects receiving the latter first. Preliminary trials revealed, however, that the difficulty of the task in Condition B when it came first was so great that many subjects, especially those in the higher age groups, gave up. Condition A was given first, therefore, to all subjects.

There were ten subjects in each of the four age decades from the twenties to the fifties. They came from the Trumpington Industrial Hostel, Cambridge, and were of various European nationalities.

When the distributions of all time measurements falling within each age range were inspected separately for each constituent reaction, it became obvious that the amount of overlap between the age groups was very large in each case, and that the variance tended to increase, often very markedly, with age, especially in respect of the scores for Condition B. All these distributions were of the asymmetrical type, positively skewed to various degrees and few of them could be described adequately by the arithmetic mean and standard deviation. In common with many other time scores, however, the logarithms of the scores were distributed approximately normally, but even so the variances remained different.

The *individual* distributions of time scores, as opposed to the *age range* distributions described above, displayed wide variations from subject to subject. It was difficult to

see what measure other than the arithmetic mean would be more representative of the central tendency of these varied distributions. Accordingly, although both geometric and arithmetic mean values were computed for each age group, the significance of observed differences was assessed from the arithmetic mean values per subject, using the method developed by Kendall (1948) and Whitfield (1947).

III

Results

In treating the results no distinction was made between the two hands since the scores for each followed the same patterning of constituent reactions studied. The separate measurements were combined, and the mean value per hand was used for the purpose of further descriptions and analysis.

Condition A

TABLE I

Mean Times (in seconds) per Subject and Differences between Age Groups. Condition A.

(Geometric mean values given in brackets).

Age group	*Initiation of response*	*Outward transit*	*Search*	*From first contact with "bull's eye" to beginning of return movement*	*Return*	*Total time*
Twenties ..	0·75 (0·73)	0·99 (0·96)	0	0·47 (0·44)	0·93 (0·90)	3·14
Thirties ..	0·85 (0·83)	1·11 (1·07)	0	0·73 (0·68)	0·96 (0·92)	3·65
Forties ..	1·03 (1·00)	1·07 (1·06)	0	0·57 (0·55)	0·93 (0·90)	3·60
Fifties	1·06 (1·01)	1·27 (1·08)	0	0·91 (0·88)	1·00 (0·95)	4·24
	For differences beetween age groups					
Subjects ranked by age ..	$P<0·01$	$P>0·1$	—	$P<0·01$	$P>0·1$	$P<0·01$
Dichotomy: Twenties/Fifties:	—	$P<0·06$	—	—	$P>0·1$	—

The total times per target together with the times for the several constituents of the performance are summarized in Table I. Four points may be noted about these results:—

1. On the average the performance of each of the four age groups was built up in the following manner: About one quarter of the total time was spent in initiating the response, about one third in moving the stylus to the target, about one sixth at the bull's eye, and one quarter in bringing the stylus back to the starting point. It is interesting to note that the outward transit time tended to be longer than the return time, which in respect of movement was essentially its equivalent.

2. The mean overall time per target for the oldest age group was about 35 per cent. higher than for the youngest. The constituent times contributing to this

increase with age were the initiation of response and the bull's eye time. The outward transit time showed a tendency in the same direction, whilst the return time remained essentially the same for all the age groups.

3. All subjects made full use of the opportunity to see while aiming at the targets—hence the complete absence of search time scores. Subjects' introspections suggested that seeing the stylus on the bull's eye rather than the restoration of the absent light in the display served to indicate success. Moreover, the experimenter observed that, upon making contact with the bull's eye, some subjects would immediately bring the stylus back to the starting point, while others would do so only after they had taken a swift glance at the display. Unfortunately, it was not possible to make this observation in such a way as to be certain whether or not it had any relation to age.

The differences in the values of initiation of response time minus outward transit time and of outward transit time minus return time gave probabilities of 0·001 and 0·01 respectively.

Condition B

TABLE II

MEAN TIMES (IN SECONDS) PER SUBJECT AND DIFFERENCES BETWEEN AGE GROUPS. CONDITION B.

(Geometric mean values given in brackets).

Age group	*Initiation of response*	*Outward transit*	*Search*	*From first contact with "bull's-eye" to beginning of return movement*	*Return*	*Total time*
Twenties ..	0·96 (0·92)	1·36 (1·28)	16·28 (7·75)	0·97 (0·88)	1·20 (1·12)	20·77
Thirties ..	1·13 (1·01)	1·29 (1·21)	17·15 (8·95)	1·06 (0·89)	1·38 (1·28)	22·01
Forties.. ..	1·55 (1·36)	1·21 (1·08)	22·58 (11·27)	1·63 (1·30)	1·58 (1·26)	28·55
Fifties	1·67 (1·50)	1·17 (1·11)	35·96 (19·08)	2·49 (1·73)	1·60 (1·36)	42·89
		For differences between age groups				
Subjects ranked by age ..	$P<0·001$	$P>0·1$	$P<0·001$	$P<0·001$	$P<0·1$	$P<0·001$
Dichotomy: Twenties/Fifties:	—	$P>0·1$	—	—	$P<0·05$	—

The results for Condition B are set out in Table II. Comparing the results in Condition B with those in Condition A, it will be at once clear that performance in Condition B took a much longer time and showed substantially greater changes with age. The main points are:—

1. The search time scores which were absent in Condition A assumed here a preponderant place, constituting some four fifths of the total time.

2. The mean overall time per target increased substantially and consistently with age, making the oldest subjects' scores twice that of the youngest. However, the performance was differently built up in the different age groups.

3. Those in the forties required on the average about 60 per cent. more time than those in the twenties to initiate the response. Those in the fifties showed a further increase of some 10 per cent.

4. No well defined differences between the groups were detected in the outward transit time. What mean trend there was, however, was in the downward direction with age. Although not significant in itself, this trend appeared of some importance when viewed against the initiation of response time. Relative to this, the youngest age group spent some 40 per cent. *more* time in the outward transit. Those in the thirties already displayed a slight shift in the opposite direction, taking only 14 per cent. more time. However, those in the forties required for the outward transit some 20 per cent. *less* time than for the initiation of the response, and those in the fifties a further 10 per cent. less still.

It should be borne in mind that the task was such that after the signal had been presented to the subject, he could interpret it at any time he chose between the moment he perceived the signal and the moment he made the first impact with the target area. This part of the performance embraced the first two constituents, the initiation of response and the outward transit. The combined scores indicate a slight increase with age, amounting to some 20 per cent. between the youngest and the oldest group.

5. The time required to locate the centre of the target increased very substantially with age. Those in the forties spent about 40 per cent. more time than the youngest on this, and those in the fifties a further 60 per cent. The time required was thus doubled from the youngest to the oldest age group.

6. It is not meaningful to give the times spent in the bull's eye for Condition B. The difficulty arises because it often happened that subjects entered the bull's eye, but went on searching apparently unaware that the display light had been restored. The mean percentage of such occurrences is given in Table III, showing a considerable increase in the forties and fifties. In all these cases of "accidental entry," the attempt to locate the bull's eye continued until at some later stage it was realized that success had been attained. Clearly, what objectively was a bull's eye was not necessarily so subjectively. The time for this stage of the performance has, therefore, been taken as the interval that elapsed between the moment an objective change in the display took place and the moment this change was apprehended by the subject so as to cause him to withdraw the stylus from the target. Those in the forties required about 55 per cent. more time than those in the twenties to become aware of and to take action after such a change in the display had taken place. Those in the fifties required some 45 per cent. more time still.

TABLE III

MEAN PERCENTAGE OF OCCURRENCES OF "UNNECESSARY SEARCH" PER SUBJECT

Age group	*Mean per hand*
Twenties ..	28·5 per cent.
Thirties ..	33·0 ,,
Forties ..	51·0 ,,
Fifties ..	74·5 ,,

7. As far as could be judged from the gross head movements, most subjects turned their heads and sometimes their bodies in the direction of the target at which they were aiming. They seemed to be attempting, although according to their introspective reports without being aware of it, to produce the same postural adjustments they would have made had they been able to see. These head movements were marked among subjects of all ages during the initial phase of the operation, and among the older subjects also during the greater part of the search. The younger subjects, however, tended during the search to watch the display—a behaviour upon which, it will be remembered, Condition B places a premium. The continuing of the attempt to locate the bull's eye after the light had reappeared was no doubt due to the fact that the subject's head was turned away from the display at the time.

8. The mean time required to bring the stylus back to the starting point showed a non-significant tendency to increase with age. The oldest group was some 20 per cent. slower than the youngest.

It is noteworthy that the mean return time, relative to the outward transit time was over 10 per cent. *lower* for those in the twenties, was almost equal for the thirties and was over 20 per cent. *higher* for the two older groups, whereas in Condition A it was lower for all age ranges.

9. At the end of the experiment subjects were encouraged to describe the subjective phenomena. When the goggles were removed, almost every subject spontaneously made remarks to the effect that the targets seemed to have been located further away than they actually were. The wording of these comments was usually such as to emphasize the subject's astonishment at how his senses had deceived him and how much nearer in reality he was to the targets. Their other remarks about Conditon B were not particularly helpful, for in this connection they referred as a rule to Condition A, in which, they claimed, they had "learnt the layout." Requested to estimate the total time spent in the experimental enclosure, most subjects overestimated by some 50 per cent. On the whole, they were puzzled and amused by the task.

One subject deserves separate mention, as showing what appears to have been an occupational transfer effect (Szafran and Welford, 1949). When he was submitted to the test, it was not known that he had had considerable experience of work in the dark-room of a photographic laboratory. His performance was for obvious reasons not included in the results of the main experiment. He was 49 years old, and by all standards his was the most efficient performance observed. Although showing the patterning of the initial phase usual for older subjects, he required less time to locate the bull's eye than the mean of the youngest age group. Also he took approximately as long as the mean of this group to react to the restoration of the display light. He practically never turned his head from the display.

With advancing age there is a decline in the efficiency with which small details can be visually discriminated (e.g. Weston, 1949). It was thought desirable, therefore, to ascertain whether the increase with age in the initiation of response time in Condition B was due to inability to see the lights or to something else.

Accordingly 18 subjects from 20 to 60 years of age were asked to lift the stylus from the starting point as soon as they noticed a light being switched off. They did not have to locate any target. Precautions were taken to ensure that they could react only to the visual signal and not to any noise produced by the switches. Every subject received ten signals for each hand in two conditions: not wearing the red goggles; wearing the red goggles.

It was found that the mean difference between these conditions lay between 15 per cent. for the younger and 25 per cent. for the older subjects. Since the percentage differences between the initiation of response times in Condition A and Condition B of the main

experiment for the corresponding age ranges were of the order of 30 per cent. for the younger and of 60 per cent. for the older, it seems clear that the differences in the initiation of response time in the main experiment were not due solely to difficulty in seeing the display.

IV

Discussion

With increase of age and transition from Condition A to Condition B there were substantial changes of performance, but they did not occur to the same extent in all the constituents of the task.

The changes with age seem to have been of three kinds.—Firstly, the older subjects tended to take longer than the younger over those portions of the task which were concerned with the *initiation of a fresh action*. Thus, in Condition A, the older tended to take longer to initiate both the outward and the return movements, although the times for the *execution* of these movements were about equal for all age groups. Likewise, in Condition B, older subjects were slower at initiating their outward movements, although the movements themselves tended to get *faster* with age. The initiation and making of the return movements in Condition B cannot properly be compared with those of Condition A, since they were likely to have been affected by the subjects' being unable to see the targets and the plate on the central pillar. The finding that older subjects took longer to initiate new responses is in line with those of a previous experiment by Brown (reported by Welford, 1951).

Secondly, the older subjects tended to be far more affected than the younger by being deprived of vision in Condition B—as is shown by their markedly longer times to locate the centres of the targets. It is possibly for the same reason that they also took longer to return from the targets to the plate on the central pillar.

Thirdly, there was a persistent tendency on the part of the older subjects to make in Condition B postural adjustments similar to those made in Condition A, when they could look at the targets while locating the bull's eyes. They were, it seemed, reinstating the postural patterns of Condition A in spite of the fact that looking away from the display made them unable to see when they had located the centre, and thus lengthened the time between hitting the bull's eye and initiating the return movement. This behaviour is interesting in view of the tendency noted in previous experiments (reported by Welford, 1951) for older subjects to look away from a display to what they are doing. Between the earlier experiments and this, there is the important difference that whereas in the former the visual information obtained *might* have been of some usefulness, in this experiment vision was excluded, and the postural adjustments must therefore have been made for some reason other than the obtaining of visual information.

What this reason is can only be conjectured at present, but two possibilities deserve consideration: First—that the postural adjustments in Condition B were patterns of action transferred from the preceding Condition A and may be regarded as due to the persistence of a habit and inability to "unlearn" a useless action—this observation has precedent in an experiment carried out by Kay (reported by Welford, 1951); second—that the proprioceptive data resulting from the postural adjustments in some way aided the older subjects in locating the centres of the targets. In other words, the older subjects required more data in order to carry out the task, and they obtained it by reinstating the postural adjustments made in Condition A.

While it is impossible definitely to choose one of these possibilities to the exclusion of the other on the data from the present experiment—it would, indeed, be very

difficult in *any* experiment—such evidence as there is seems to favour the second rather than the first. For not only were the older subjects less efficient without vision, but also they were in many cases unable to carry out the task in Condition B at all unless they had done it first in Condition A, although they inspected the display and the targets before the experiment began. Thus they illustrated what Sherrington has described as "the inefficacy, as a means to arrive at a new motor correlation, of instruction merely verbal, or of ideas constructed without motor experience." Clearly something transferred from Condition A enabled the older subjects to carry out the task in Condition B, and it seems reasonable to suppose that this was a series of proprioceptive patterns.

The experiment was originally suggested by Professor Sir Frederic Bartlett and was carried out under the general supervision of Mr. A. T. Welford. Mr. J. A. Leonard designed and constructed the electrical circuit, in consultation with Mr. N. T. Welford. The writer wishes to thank Miss V. R. Cane and Mr. A. E. D. Schonfield for some helpful criticism.

References

1. Fitts, P. M. (1947). A study of location discrimination ability. (In *U.S.A.A.F. Research Report No.* 19: *Psychological research on equipment design.* Edited by P. M. Fitts.)
2. Kendall, M. G. (1948). *Rank Correlation Methods.* London.
3. Sherrington, C. S. (1905). The Integrative Action of the Nervous System. 1947 edition. London.
4. Szafran, J., and Welford, A. T. (1949). On the problem of generalized occupational transfer effects in relation to studies of ageing. *Quart. J. exp. Psych.*, **1**, 4.
5. Welford, A. T. (1951). *Skill and Age.* Oxford
6. Weston, H. C. (1949). *Sight, Light and Efficiency.* London.
7. Whitfield, J. W. (1947). Rank corrrelation between two variables, one of which is ranked, the other dichotomous. *Biometrika*, **34**, 292–6.

(*Received 4th March*, 1951.)

THE OPERATION OF AN ACQUIRED DRIVE IN SATIATED RATS

BY

K. DANZIGER

(*From the Institute of Experimental Psychology, Oxford*)

A series of four short experiments indicates that the behaviour of satiated rats in a runway, at the end of which they have previously been rewarded, differs significantly from the behaviour of satiated rats without previous reward in the experimental situation. The former group reach the end box more quickly after having been put in the starting box of the runway and if provided with food in the end box proceed to eat it, although they have just refused similar pieces of food in their home cages. This is shown not to be due to defective satiation or the operation of fear in the control group. When runs and feeding in the end box are separated during the training period, the previously rewarded group still shows a more vigorous response on satiated trials, indicating that it is a reward expectancy about the goal box rather than a running habit which has become "functionally autonomous," acting as a situation-specific drive.

Introduction

Most experiments on animal learning make use only of the basic physiological drives because of their dependability. The result is that the animal almost never gets an opportunity to demonstrate any ability it may have to learn drives as well as perceptual and motor habits. This has led some psychologists to infer that only the basic physiological needs do in fact act as drives in sub-primates, and Allport (1947), for instance, concludes that the gap between human and animal motivation cannot be bridged.

Nevertheless, a few recent experiments have demonstrated the operation of acquired, situation-specific drives in an animal like the rat. In particular, it has been established that fear repsonses may be acquired (Estes & Skinner, 1941, May 1948, Miller, 1948b, and Mowrer, 1940); that is to say, a neutral stimulus repeatedly presented with an electric shock may eventually produce marked avoidance behaviour when presented alone. It is not possible, however, to generalize about drive acquisition from this evidence, as it is obtained from an abnormal situation highly disturbing to the animal.

Drive acquisition based on the food-seeking tendency is not nearly so well corroborated as drive acquisition based on fear. The best report is that of Young (1946) which shows that rats can acquire specific food habits which may persist in opposition to the physiological need of the animal.

The well-known work of Anderson (1941) on what he called "drive externalization" is perhaps only suggestive. Non-rewarded and satiated rats were shown to give evidence of maze learning provided they had previously been rewarded for learning another maze. But serious doubt has been cast on the efficacy of Anderson's satiation technique and on the significance of the response measure he used which may have shown up simply some transfer of training effects (Siegel, 1943).

Apart from mere restatements of observed facts the only attempt at providing a theoretical basis for the element of learning in motivation has been made by Hebb (1949). He demonstrated that even in the case of an apparently innate drive like hunger, there is a learned element when it comes to eating in a particular situation. He regards hunger as an organized neural activity with a motor facilitation of the various activities of eating and with numerous connections to other organized neural systems. From this it seems to follow that the tendency to eat may be aroused by the

activity of other systems corresponding to perceptions or symbolic activities, possibly even in the absence of whatever indications of internal need normally play a part in arousing it. If this is so, it would provide a basis for specific acquired drives of a rather situation-bound type. Anderson's "externalization" principle by which a drive at first aroused by internal factors comes, by a process of learning, to be aroused by external factors suggests much the same thing.

But there are difficulties in these conceptions. So far as the acquisition of the new element is concerned, it is not clear whether the mere presence of an unsatisfied drive in a certain situation suffices or whether it is necessary for a consummatory act to occur. It is also important to know whether the new element operates like an innate drive, activating appropriate instrumental and consummatory acts. None of these questions can be decided on the basis of existing investigations.

It was therefore decided to set up a series of experiments designed to show whether and how originally neutral stimulus-situations associated with hunger and food reward were capable of affecting the behaviour of satiated rats. For this purpose it was necessary to measure performance in a very simple situation, such as a straight runway which the animal knew well, so that cognitive learning was at a minimum and did not obscure the results. It was also important to make sure that the primary drive had been effectively satiated when the animals were tested for the existence of acquired drives.

Apparatus

The apparatus consisted simply of a runway 5 feet long and 5 inches wide with sides of plywood 1 foot high and a perforated zinc floor. At either end this runway was attached to an enclosure 9 inches × 10 inches × 12 inches; one of which was painted black and used as a goal box. During the preliminary trials, while the rats were getting used to the apparatus, the tops of the two end boxes were coverd by netting. During the experimental trials rats only very rarely attempted to jump out of the apparatus and so the netting was folded back. The runway also contained three hurdles placed at regular intervals. They consisted of blocks of wood the same width as the alley, 4 inches long and 1 inch high. They were introduced to slow up the rat's running speed so that any errors in measurement would form a less appreciable fraction of the observed value. The black end box contained a shallow white dish in which could be placed a food pellet on rewarded trials.

The whole apparatus was placed on a concrete bench in the same room as the rats' home cages. It was not possible to control illumination from day to day, but if this was a disturbing factor, which seems rather unlikely, it must have affected experimental and control animals equally. Random auditory stimulation in the animal house may have produced occasional high time scores, but hardly any consistent effect on the results of the experiments, as the animals, being used to this sort of stimulation, showed quite a high tolerance for noise.

Subjects

The rats used in these experiments all belonged to the same strain of hooded animals bred originally by the National Institute for Research in Dairying at Reading. They were all three to four months old at the start of their experimental training. Some animals were rejected in the early stages of the experiments for persistent failure to adapt to experimental conditions, shown by refusal to run, squealing, jumping, defecation, etc.

Control of Feeding and Satiation

For a fortnight before the first preliminary trial the rats were all handled daily by the experimenter in order to tame them. During the last few days of this period their food in the form of dry rat-cake nuts was supplied only once a day for 90 minutes, at the end of which period it was removed. In this way a 24-hour feeding schedule was established, and this was maintained throughout the experiments. It ensured a constant and maximum degree of hunger immediately before the daily feeding period and a pretty complete state of satiation at the end of that period. Tests made during preliminary experiments had

indicated that this method of satiation was as effective as more tedious and elaborate methods. Rats would always stop eating spontaneously after 60–80 minutes, so that the food was never removed when an animal was still eating. Water was available in the home cages at all times.

The food pellets used as an incentive on rewarded trials consisted of the same composite rat food that made up the rats' daily food ration. They were prepared by adding starch to a wet mash of rat-cake nuts, rolling the resulting dough, cutting it into even pellets and then drying these in an air oven. The weight of the pellets was in the neighbourhood of 0·22 gm., varying slightly, but never by more than 0.04 gm. Thus a particular rat's reward might fluctuate by almost 50 per cent. from one day to another. However, as an investigation by Miller & Lawrence (1947) seems to show that it needs at least a fourfold difference in the amount of food reward to bring about a significant change in the rat's behaviour, these variations may be neglected.

Experiment I

The object of this experiment was to determine whether a simple instrumental act, like running to a goal-box would be performed more vigorously by satiated rats with previous experience of hunger and reward in this situation than by satiated rats having the same acquaintance with the runway, but never having experienced hunger or reward in it. Such a difference in performance would have to be put down to a difference in motivation rather than in ability, and with the internal innate elements at a minimum for both groups the difference would indicate the effectiveness of learned elements in the motivation of the first group.

Procedure

After a period of preliminary handling each of the 12 male rats used in the experiment was left in the apparatus for a period of eight minutes after it had received its daily food ration. On the succeeding three days the rats were left in the apparatus for a three-minute period after receiving their daily food ration. The novelty of the situation ensured very active exploration of the runway, even though the animals were not hungry. In this way the rats were familiarized with the experimental situation and their initial fear response was greatly reduced, but there was as yet no association of either hunger or food with the experimental situation.

On the fifth day of the experiment the animals were divided into two groups of six rats each. The experimental group were given one rewarded run daily 15-30 minutes before their daily food ration was due until a total of 12 such runs had been made. On the succeeding five days they were given one run daily to an empty end box immediately after having received their daily food ration. The control group was given 17 successive daily runs to an empty goal box immediately after satiation. The crucial comparison is between the performance of the two groups on the last five trials.

The method of measuring performance at each trial was as follows. At the beginning of each trial the rat was dropped into the starting box facing the wall opposite the alley entrance, so that in order to run down the alley it had to turn about. As the rat touched the floor of the starting box the experimenter started a stopwatch equipped with two hands and stepped back behind the starting box to a position in which he was invisible to the rat but from which he could just observe the entrance to the alley. When both the rat's forelegs were actually in the alley and out of the starting box the first hand of the watch was stopped. The measure obtained was therefore one of the speed with which the rat reacted to the starting box by setting off on its run down the alley. When the rat had three feet inside the end box, the second hand of the watch was stopped so that the difference between the first and the second readings gave a measure of the time taken by the rat to traverse the 5-foot alley. On reaching the goal box after each trial the rat was placed in a small separate cage for two or three minutes before being returned to its home cage. Time readings were taken to an accuracy of 0·1 second.

Retracing was not prevented, because the retracing animal is obviously poorly motivated to reach the other end of the runway, and this should show in his time score. The occasional very high time scores produced in this way can be prevented from weighting the results too heavily by converting all scores to a logarithmic scale before comparisons are made. This is necessary for time scores in any case and is the usual procedure for dealing with this type of data. Its statistical soundness is well established (Gaddum, 1945).

It may be thought that the latency of response measure would not be reliable because of slight variations in the way the animals were dropped into the starting box, but in fact it was found that this measure was less subject to random fluctuations than the running speed.

On the first rewarded trial many of the animals in the experimental group ate the reward spontaneously, while the others could be readily coaxed to do so by offering the reward pellet by hand in the end box. On all succeeding trials consumption of the reward was immediate and spontaneous. These observations indicate that the four days of preliminary training were rather effective in reducing the initial fear response to the experimental situation.

Results

Running time in the alley and response latency in the starting-box were summed for each trial to give the total response time which provided a good composite measure of performance. This was converted to logarithmic form, and the mean log total response time was calculated for each trial. During the first 12 trials the performance of the experimental group improves progressively, but on the 13th trial which is the first satiated trial, there is a sharp increase in time, though not nearly to the response level of the control group. Table I compares the performance of experimentals and controls on the last eight days of the experiment. On trials 10, 11, 7, 12, the experimentals were still running hungry and rewarded. Trials 13–17 are the test trials on which both groups are run satiated and without reward.

TABLE I

1,000 X Mean Log Total Response Time on Last Eight Trials

	Trials							
	10	11	12	13	14	15	16	17
Experimentals	653	584	581	940	657	800	888	990
Controls	1,257	1,141	1,140	1,228	1,233	1,202	1,105	1,333

In testing for the significance of the difference between groups on test trials we have to take into account that individual performances, by control animals in particular, are subject to random variation over a considerable range. This leads to very high standard error estimates which reduce the significance of the difference between trial means. We can, however, make an overall comparison of the two groups on all the five test trials, by taking the average performance of each rat for all the five test trials and then comparing the values so obtained for the two groups. In this way day-to-day fluctuations will be minimized for each animal and a much better distribution on which to do the test will be obtained.

TABLE II

1,000 X Mean Log Total Response Time of All Five Test Trials for Each Rat

	Rats						
	1	2	3	4	5	6	*Overall mean*
Experimentals	880	776	750	851	862	1,010	855
Controls	1,296	1,496	1,434	1,290	732	1,074	1,220

It will be seen from Table II that with one exception the rats in the experimental group are all responding more quickly (i.e. make lower time scores) than those in the control group. The exception is one rat in the control group which responds faster than any other rat in any group. Notes on this animal show it to have been very active not only in the experimental situation. This type of gross individual difference does occur occasionally in work of this kind, especially if the strain of experimental animals is not highly inbred.

Mean log values for the running speed and reaction latency of each rat were calculated, giving a set of pairs of overall mean values which can be compared by the rest. (*See* Table III.)

TABLE III

Comparison of Average Values of Performance over Five Test Trials

	Total response time		*Running time*		*Reaction latency*	
	Exp.	*Contr.*	*Exp.*	*Contr.*	*Exp.*	*Contr.*
Mean ..	855	1,220	659	1,022	338	647
t ..	2·78		2·74		3·14	
P ..	0·02		0·02		0·01	

For 10 degrees of freedom the 1 per cent. level of confidence is actually reached by a *t* value of 3·17 and the 2 per cent. level by a *t* value of 2·76.

These results seem to indicate that satiated rats which have previously received food in the end box of an alley will reach this end box more quickly, when given an opportunity, than satiated rats which have never received food there or even been hungry while in the alley. This difference affects both the speed with which the rats actually run the alley and the time they spend in the starting box before starting on their run.

Discussion

The rats in the experimental group appear to be motivated to reach the end box, although presumably, they have no need of the food which they have learned to expect there. One could say that the "reward expectancy" built up in the runway on rewarded trials becomes "functionally autonomous" by operating even when basic physiological need is in abeyance.

Comparable results have been obtained recently in a number of experimental situations. Zeaman & House (1950) have shown that a habit that has led to the satisfaction of a light-avoiding drive will persist at quite a high level when the original drive is entirely eliminated. This result is especially interesting because a primary drive depending entirely on external factors can, of course, be far more convincingly eliminated than a vegetative drive like hunger. Using hunger, Miller (1948) reports almost errorless performance on first trials by satiated rats in a maze in which they had previously been rewarded. Webb (1949) found incidentally that a group of rats satiated for both food and water would still respond to a door behind which they had previously found food pellets.

The question immediately arises of whether we are not in all these cases dealing with examples of habit persistence; acts which have been repeated so often in a certain situation that they run off mechanically even in the absence of the original motivation. Hull has suggested that any drive may activate any habit, so that in the present

experiment the strong running habit of the experimental group may be activated by another, "irrelevant," primary drive, even though the hunger drive has been satiated. This effect would not be observed in the control group, which has had no previous reinforced trials and so no habit that can be activated. In order to decide between this explanation and one in terms of learned elements in motivation further experiments are necessary. To begin with the motivation hypothesis would lead us to expect an observable effect on consummatory as well as on instrumental behaviour. If the experimental situation does indeed regenerate an eating drive in the satiated rat, then it should show some tendency to satisfy this drive apart from running to the food box; it should in fact eat when confronted with food there. In order to test this the second experiment was performed.

Experiment II

The subsidiary purpose of this experiment was to determine whether an external situation acquired the property of evoking a secondary drive because of its previous association with the primary drive state or whether a consummatory act in the situation was also necessary. This problem is of some importance from the point of view of systematic learning theory, as the rival principles of "contiguity" and "effect" are involved.

Procedure

The apparatus and general method of conducting each trial were the same as in Experiment I. Ten male rats of the same strain were used. The actual schedule for the experiment was as follows. On the first four days of the experiment all 10 rats were given daily 10-minute periods of free exploration in the runway after receiving their daily food ration. In this way they were made familiar with the apparatus, without any association of either hunger or food with the experimental situation.

On the fifth day of the experiment the animals were divided into an experimental and a control group of five rats each. The experimentals were given one rewarded run daily for the next 12 days, while the controls were run non-rewarded for the same number of trials. All runs took place about half an hour before the daily feeding period. On the succeeding five days both groups were given one run daily immediately after the feeding period, a reward pellet being provided in the end box for both groups. If the animals did not eat this pellet within one minute of reaching the goal box, they were taken out of the apparatus. Running times were taken for both groups.

Results and Discussion

Of the five animals in the experimental group, four consumed the pellet of food provided in the end box on all five satiated trials, within a few seconds of reaching the end box. The fifth rat in this group showed the same behaviour on the first two of the satiated trials, but did not eat the pellet within one minute of reaching the end box on the last three trials. Of the five animals in the control group, which had never eaten in the end box, four never touched the food pellet on satiated trials, the fifth consumed the pellet on one of the five trials. These results are not amenable to statistical treatment, because the number of subjects is so small, but they are certainly suggestive.

Not only did the two groups show differences in their reaction to the food provided in the end box, they also differed with respect to the speed with which they approached the end box as in Experiment I. The running times obtained on satiated trials were converted to their logarithms and the performance over the five trials was then averaged for each rat. It was found that the difference between the mean running times of experimentals and controls was significant at the 5 per cent. level.

Thus the behaviour of the experimental group appears to show evidence of the continued action of some sort of hunger drive after satiation in the home cage, while the behaviour of the control group does not. This drive seems to show itself both in the tendency for a consummatory reaction to take place in the end box and in the increased speed in getting there. The previous rewarded runs seem to have led to an "externalization" of drive, whereby the cues from the apparatus became able to arouse a certain residue of drive even in the absence of organic need. Also it seems that this effect definitely depends on previous experience of reward in the situation to which the drive is "externalized"; the mere presence of unsatisfied drive, as in the control group, has no such effect.

It is apparent that the significance of the present experiment depends on the assumption that the primary drive is at a minimum immediately after cessation of the consummatory act of eating. It is important to remember that we are dealing here with the behavioural effects of the drive and that these may not be directly correlated with the retoration of physiological equilibrium. Behavioural studies do indeed show quite clearly that primary drive action is at a minimum immediately after the completion of consummatory activity and not at some time afterwards. The best empirical study is that of Koch and Saltzman (1948); Zeaman & House (1950) in a recent discussion of the general problem of the zero point for primary drives quote a number of references.

But the results of this experiment can hardly carry conviction as they stand. Quite apart from the small numbers involved, everything depends on the efficiency of the satiation technique employed in really eliminating any general need for food. If both groups were under some residual drive when starting their test runs then the differences in their behaviour could be accounted for by the fact that the control group had never previously fed in the apparatus and so would have to be under a strong need to "take the plunge" and eat in this unusual situation. Moreover, their reluctance to eat would be enhanced by the fact that they had had no previous experience with the pellets provided in the end box. It is thus necessary to check up on three points: the satiation technique, the effect of fear on failure to eat in a relatively unfamiliar situation and the effect of strangeness of the reward object on consummatory behaviour.

The last point was readily cleared up. The control group of six rats used in Experiment I had been given 17 non-rewarded runs in the apparatus on the same days as the animals used in the present experiment. On the 18th day they were fed by hand liberal quantities of reward pellets in their home cages with their daily food ration. After satiation they were run in the apparatus and provided with a pellet in the end box. None of them consumed this pellet within one minute. As a further test the pellet was presented to them by hand for another two minutes while they were in the end box. There was much sniffing but no eating of the pellet. It appears then, that the results of the first experiment were not due to the lack of familiarity with the reward object on the part of the control animals.

Experiment III

This experiment was designed to determine whether the eating behaviour shown by satiated rats in a place where they have previously been regularly rewarded is due to a defective satiation technique, and whether the failure of satiated rats to eat in a new situation is due to fear of eating in a place in which they have never eaten before.

The satiation technique used consisted in daily feeding periods on rat-cake, but on the test day, each rat, after having finished eating this, was fed experimental

reward pellets by hand until it repeatedly rejected them. Then it was immediately given its test run with a pellet in the end box. If it ate this pellet, when it had just refused an exactly similar one in the home cage, this could hardly have been due to defective satiation.

Most rats are shy about eating in an unfamiliar situation. But when they have once been coaxed into eating in a place, they will always do so again spontaneously on the next occasion. Hall (1936) has shown that a sharp drop in emotional behaviour occurs when animals have once eaten in an unfamiliar situation. The control group in the present experiment was therefore given a single rewarded trial when hungry, a few hours before the test trial. Their single experience of eating in the end box should have removed their inhibition about doing so, and so if they still did not eat on the satiated trial, this must be simply due to the effectiveness of the satiation procedure. But a single eating experience in the end box must have begun the process of "externalization," and so the results of this group would probably not be quite as negative as those of the control groups in the first experiment.

Apparatus, details of procedure, preliminary training, feeding schedule and reward pellets were the same as in the first two experiments. The subjects were 17 male rats, 9 experimentals and 8 controls, of the same strain as those used in the previous experiment and about four months old at the start of the experiment. After four days of ten-minute exploration periods in the runway, all the animals were given three runs a day for the next three days. This method was adopted to save time. Both groups were run 15–30 minutes before their daily feeding period; the experimentals found a food pellet in the end box, and the controls were removed from the empty end box after 30 seconds. There was a five-minute interval between each of the daily trials. This was spent in a small, special cage kept for this purpose. On the next day, the test day, the procedure was the same for both groups. There was one rewarded trial when hungry, then the satiation procedure already described, and finally the test run immediately afterwards. The animals were given one minute to consume the pellet provided on this run, and if they had not begun eating at the end of this period, they were gently offered the pellet by hand while in the end box for another two minutes. At the time of the test-run the experimental group had therefore had 10 previous rewarded trials, while the control group had had nine non-rewarded and once rewarded trial.

Results and Discussion

All the nine rats in the experimental group consumed the pellet on the test run, but only three out of eight in the control group did so. These results can be drawn up in a 2×2 table and treated by Fisher's exact method. This gives a probability just less than 0·01. Applying Yate's correction the table also yields X^2 value of 5·24 which gives a P value just above 0·02. The difference in the eating behaviour of the groups may therefore be taken as statistically significant.

Testing the difference in the mean running speed of the two groups on the test trial, a value of 2·85 is obtained, which at 15 d.f. is significant at the 2 per cent. level.

The results of Experiment II both with regard to the consummatory reaction and the instrumental running activity have therefore been confirmed. That rats which have just rejected a food pellet in their home cage will run much quicker than satiated rats normally do to a box where they know another such pellet to be and then eat this pellet suggests that the eating drive can be aroused by specific situations. It also suggests that the effects of such situations are surprisingly independent of one another, so that satiation in one does not make a fresh arousal in another situation impossible.

The fact that as many as three out of eight animals did eat on the test trial is perhaps surprising. It will be remembered that all eight rats had eaten in the experimental situation once before, so that there could be no question of a feeding inhibition

due to lack of familiarity. We can only conclude that either the first rewarded trial is rather more important than any succeeding ones in setting up "externalization" effects, or else that "externalization" needs only very few trials altogether. This can only be tested by further experiment. General considerations about the shape of learning curves and the rapid learning of determining sets at the beginning of the curve make both these possibilities likely ones, however.

A number of investigators have reported eating or drinking by rats satiated for food or water. In considering their results it is important to remember that eating and drinking like any other consummatory act, may appear as "displacement activities" (Tinbergen, 1940), when another drive is strongly aroused while its own consummatory act is not possible in the situation. Eating or drinking behaviour in satiated rats can only be taken as evidence for the operation of an acquired situation—specific drive if there are no other strong drives operating at the time. An experiment performed by Miller (1948) provides some interesting data on this point. In order to demonstrate displacement drinking two groups of rats were trained to run to a water reward in a runway similar to the one used in the present investigation. One group was then run satiated for both food and water and another group was run hungry but not thirsty. There was some drinking at the end of the runway by the first group, but much more by the second. The difference between the amounts drunk by the two groups is due to the displacement factor, but this does not account for the drinking by the first group, which was presumably due to a situation-specific thirst drive. Drew (1937) has also shown that satiated rats will eat (*a*) in a place in which they have previously eaten, and (*b*) in certain situations which seem to involve conflict for the animal. By the present interpretation we get eating due to a situation-specific eating drive in the first case and displacement eating in the second.

Experiment IV

Introduction

In all three experiments reported so far the behaviour of the rats on satiated trials has been very similar to the behaviour they displayed on previous rewarded trials. The rats in the experimental groups have undoubtedly learned something which the controls have not, and this results in their different performance on satiated trials. But what is it that has been acquired? Is it a tendency to repeat mechanically all the actions of the rewarded trials irrespective of primary motivation, or is it some dynamic set specific to the situation?

In order to answer this question, the preliminary running to the end box of the apparatus was separated from the obtaining of reward there. The experimental animals were at no time run to an end box containing food, so that if they were to run faster on satiated trials, this could not be due to the persistence of the running habit. Nor could it be due to the sensitizing of a habit by an irrelevant drive, because no opportunity was given for any habit to form. If one now gives food to the satiated rats in the box in which they have previously been fed, will they eat it? If they do, will they still run to that box faster when given an opportunity immediately afterwards?

Procedure

The apparatus and control of feeding and satiation were the same as in Experiment I. Fourteen female rats of the same strain were used as subjects. Running time and latency of response were measured in the same way as before.

On each of the first four days of the experiment the animals were allowed to explore the apparatus freely for five minutes, a short while before their daily feeding period. On the

next five days they were each given a single non-rewarded run daily when hungry. The purpose of this nine-day preliminary training period was to allow the animals to familiarize themselves thoroughly with the runway. On the 10th day the animals were divided into two groups of seven rats each. The end box was shut off from the rest of the apparatus by a sliding door. About an hour before their daily food ration was due each of the 14 rats was placed directly in the end box and allowed to remain there for two minutes. About 20 minutes later they were put into the end box for another 40-second period, and 20 minutes after that for a final period of 40 seconds. A little while later they received the daily food supply in their home cages. The only difference between experimental and control groups was that the former were allowed to consume a quantity of food in the end box while for the latter the end box was empty. Thus the end box was associated with reward for the experimental group, but that reward had never been preceded by a run down the alley. At the end of the usual daily feeding period each of the animals was once more put in the closed end box for a period of 40 seconds. On this occasion too, the end box was empty for the controls, but contained food for the experimentals. At the end of the 40-second period the animal was lifted out, the sliding door was removed, the animal was taken to the starting box and was allowed to run down the alley in the usual way, its time being measured. The run led to an empty end box for both groups.

On the 11th, 12th, 13th and 14th days of the experiment the procedure was to give each animal a 40-second period in the closed end box about 30 minutes before the daily food supply, and then to give them one run immediately after satiation, without any further stay in the end box. The end box was always empty for control animals, but contained food when experimental animals were introduced for their 40-second stay. It never contained food on satiated runs.

Results

As in the first experiment the two response measures, running times and latency of reaction were converted to their logarithms and mean log. values were calculated for each trial. Table IV gives these mean values for the running speed on the last six days of the experiment. As in Experiment I, log. values are multiplied by 1,000 to eliminate decimals.

TABLE IV

1,000 X Mean Log Running Time on Last Six Trials

	Days							
	9	10	11	12	13	14	5–9	11–14
Experimentals	1,158	1,091	760	1,013	972	918	981	917
Controls	1,084	981	1,255	1,148	1,334	1,194	1,044	1,233

The ninth day, is the last day on which both groups are run hungry and without reward. It marks the end of the preliminary period of familiarization. On all subsequent days the animals are run satiated, but the experimentals have been receiving short periods of feeding in the end box. The procedure on day 10 is different to that on other days in that a period in the end box immediately precedes the satiated run. For the experimental group the end box contained a pellet of food on this occasion as usual. Now, although they had just refused any more food in their home cages, all seven animals in this group proceeded to eat the pellet available in the end box. It will be seen that their subsequent run was rather slower than any later run, when this extra pellet was not provided. Using the method for correlated samples, the difference in performance of the experimental group on days 10 and 11 yields a t value of 6·04 which makes $P < 0{\cdot}001$. The difference between the two days is therefore

not likely to be due to chance fluctuation. The last two columns in the table give the average values for the five preliminary hungry non-rewarded trials and for the last four satiated trials for each group. The difference between running speed during hungry and during satiated trials are significant for neither group.

TABLE V

SIGNIFICANCE OF DIFFERENCES BETWEEN MEAN LOG RUNNING TIMES OF EXPERIMENTAL AND CONTROL GROUPS ON VARIOUS DAYS AND GROUPS OF DAYS

	Days					
	9	10	11	5–9	11–14	12–14
t ..	0·287	1·76	2·73	0·073	2·64	1·99
P ..	<0·8	0·1^{+}	<0·02	>0·9	0·02^{+}	<0·1

For 12 degrees of freedom a *P* value of 0·1 actually requires a *t* value of 1·78 and a *P* value of 0·02 requires a *t* value of 2·68.

The comparison between the average performance on days 5–9 and on day 9 separately shows that there was no significant difference in the performance of the groups before they were subjected to different experimental treatment. It is to be noted that the difference on day 10 is in the opposite direction to those on the following days, i.e. on this day the controls are faster.

Similar comparisons using reaction latencies instead of running times give results that lie in the same direction, but are less significant statistically. However, the difference between reaction latencies of experimentals and controls on day 11 is still significant at the 5% level.

What must be explained, therefore, is the fast performance of controls and the retardation of experimentals on the first satiated trial, the extremely fast performance of experimentals on the second satiated trial, and the persistent rather faster performance of experimentals on subsequent satiated trials.

DISCUSSION

A possible interpretation of the difference between the groups on days 11–14 is that the periods of confinement in the closed, empty end box strengthened any fear reaction which the control animals may have had initially to this part of the apparatus. This would cause them to approach the end box more slowly on satiated trials than the experimental animals. The activity level of satiated rats is, however, well below that of hungry rats, and an unpublished experiment of the writer's showed that hungry rats will always run faster in this type of apparatus than satiated rats, even though neither group receives a reward. What must be explained in the present experiment, therefore, is not why the controls tend to run more slowly on satiated trials, but why the experimentals fail to do so. In any case, if the control animals had acquired a fear reaction to the end box, this would show in their score on the first satiated trial (day 10), because the run was only taken as completed if the animals had three legs actually inside the end box. But in fact five of the seven animals in this group actually entered the end box more quickly on day 10 than on day 9 before they had ever been confined there.

So far as the main object of the experiment is concerned, there is no evidence for the formation of a "running habit" on days 5–9. The mean time score on day 9 is actually well above the mean for all five days, so that if anything, the animals are slowing up as they get completely familiarized with the runway. Their tendency to run to the end box much more quickly than the controls on day 11 and rather more quickly on the following days is therefore not due to the persistence of a running habit, whether sensitized by an irrelevant drive or running off mechanically. A possible explanation of the faster running times of experimental animals on satiated trials might be given along the following lines. The animals presumably have some memory of feeding in the end box, which is probably related to an expectation of feeding there again. Now what seems to happen is that this expectation acts like a motive in that it activates behaviour which is likely to lead to its satisfaction. In the present setting the animal makes use of previously acquired information about the runway in an attempt to bring about the satisfaction of its expectation of being fed in the end box. That this can occur in a hungry animal has already been shown by Seward (1949), who first allowed rats to explore a single unit T-maze thoroughly, then fed them in one end box and immediately afterwards gave them a free run from the starting box. Almost all the animals ran to the side on which they had just received food. Seward's animals were still hungry when making their choice, but in the present experiment the reward expectation seems to be operating independently of the primary drive. This too, is not entirely unknown, for MacCorquodale & Meehl (1949) found that rats satiated for both food and water would gradually prefer that side of a single unit T-maze on which these incentives were provided.

The effect we have been discussing only appears in pure form on the 11th day of the experiment. On subsequent trials it is much less well marked because the animals are presumably acquiring a discriminative expectation as a result of being repeatedly *not* fed when reaching the end box at the end of a run. Presumably, whereas before the expectation was simply, "food in end box," they are now beginning to distinguish between "food when put in end box," and "no food when running to end box." The fact that differences between groups are less well marked on reaction latencies than on running times is probably due to the difficulty which the rats have in distinguishing the starting box from the end box. Before they learn to react immediately to the brightness difference, the experimental rats probably spend some time looking for food in the starting box and this will increase their latency score.

The results of the 10th day must now be considered. On that day the experimental animals all ate the pellet of food provided in the end box, even though they were supposedly satiated, and showed a rather retarded performance immediately afterwards. The eating behaviour is in line with the results of Experiments II and III, and incidentally, it shows that it does not depend on a previous run which may have built up an energy debt. It is apparent that the end box instigates eating behaviour irrespective of what eating may have taken place elsewhere. It is also clear that those situations which are capable of instigating eating behaviour are situations in which the animal has fed previously; they are the situations too, in which, as we phrased it above, it expects to feed again. In other words, expecting to eat and wanting to eat are equivalent here, and a "reward expectancy" seems to be acting like a highly specific motive. The wider theoretical implications of these views will be discussed elsewhere (Danziger, 1951).

What happens on day 10, therefore, is that the specific reward expectancy about feeding in the end box is satisfied before the run and so cannot act as a motive. The animals therefore run, if anything, more slowly than the controls. It is possible that

there may actually be an inhibitory effect involved, but this will require more experimentation. In any case, the views put forward in this discussion are at present only tentative and await further experimental exploration and confirmation.

Conclusions

The behaviour of satiated rats seems to indicate that when rats have had some experience of eating in a certain situation, the reward expectancy connected with that situation is able to operate in the absence of any physiological hunger drive in such a way that the animals display the sort of behaviour that would lead to the satisfaction of the reward expectancy in a constant environment.

Thus when a tendency to eat is aroused by the external situation it acts like an internally aroused drive in that it activates learned instrumental acts and utilizes stored information. Satiation of a primary drive in one situation does not make it impossible for a different situation to arouse this drive again. But only situations associated with a consummatory act and not those associated with an unsatisfied need state acquire the capacity to rearouse the drive in the satiated animal. This suggests that the process may be one of a reward expectancy becoming "functionally autonomous."

The consummatory action appropriate to the primary drive is important for the continued operation of the secondary drive. In this way the acquired drives of rats seem to differ from those of men.

I wish to express my thanks to Professor Humphrey and Dr. H. M. Sinclair for supplying the research facilities and to Mr. B. M. Foss for his advice and encouragement.

References

1. Allport, G. W. (1947). Scientific models and human morals. *Psychol. Rev.*, **54**, 182–192.
2. Anderson, E. E. (1941a). The externalization of drive. I.—Theoretical considerations. *Psychol. Rev.*, **48**, 204–224.
3. Anderson, E. E. (1941b). The externalization of drive. III.—Maze learning by non-rewarded and by satiated rats. *J. Genet. Psychol.*, **59**, 387–426.
4. Danziger, K. (1951). Motivation in the light of the matching hypothesis. To be published.
5. Drew, G. C. (1937). The recurrence of eating in rats after apparent satiation. *Proc. Zool. Soc.*, Ser. A, Part I, 95–106.
6. Estes, W., and Skinner, B. F. (1941). Some quantitative properties of anxiety. *J. exp. Psychol.*, **29**, 390–400.
7. R. A. Fisher. *Statistical Methods for Research Workers.* 10th Ed. Edinburgh.
8. Gaddum, J. H. (1945). Log normal distributions. *Nature*, **156**, 463–466.
9. Hall, C. S. (1936). Relationship between emotionally and ambulatory activity. *J. comp. Psychol.*, **22**, 345–352.
10. Hebb, D. O. (1949). *The Organization of Behaviour.* New York.
11. Hull, C. L. (1943). *Principles of Behaviour.* New York.
12. Koch, S., and Saltzman, I. (1948). The effect of low intensities of hunger on the behaviour mediated by a habit of maximum strength. *J. exp. Psychol.*, **38**, 347–370.
13. MacCorquodale, K., and Meehl, P. E. (1949). Cognitive learning in the absence of competition of incentive. *J. Comp. Phys. Psychol.*, **42**, 383–390.
14. May, M.A. (1948). Experimentally acquired drives. *J. exp. Psychol.*, **38**, 66–77.
15. Miller, N. E. (1947). Experiments on the strength of acquired drive based on hunger. *Amer. Psychologist*, **2**, 303.
16. Miller, N. E. (1948a). Studies of fear as an acquirable drive.—I. *J. exp. Psychol.*, **38**, 89–101.

17. Miller, N. E. (1948b). Theory and experiment relating psycho-analytic displacement to stimulus-response generalization. *J. Abnor. & Soc. Psychol.*, **43**, 155–178.
18. Miller, N. E., and Lawrence, D. H. (1947). A positive relationship between reinforcement and resistance to extinction by removing a source of confusion from a technique that had produced opposite results. J. *exp. Psychol.*, **37**, 494–509.
19. Mowrer, O. H. (1940). Anxiety-reduction and learning. *J. exp. Psychol.*, **27**, 497–516.
20. Seward, J. P. (1949). An experimental analysis of latent learning. *J. exp. Psychol.*, **39**, 177–186.
21. Siegel, P. S. (1943). Drive-shift, a conceptional and experimental analysis. *J. comp. Psychol.*, **35**, 139–148.
22. Tinbergen, N. (1940). Die Uebersprungbewergung. *Z. Tierpsychol.* **4**, 1–40.
23. Webb, W. B. (1949). The motivational aspect of an irrelevant drive in the behaviour of the white rat. *J. exp. Psychol.*, **39**, 1–14.
24. Young, P. T. (1946). Studies in food preference, appetite and dietary habit. VI.—Habit, palatability and the diet as factors regulating the selection of food by the rat. *J. comp. Psychol.*, **39**, 139–176.
25. Zeaman, D., and House, B. J. (1950). Response latency at zero drive after varying numbers of reinforcements. *J. exp. Psychol.*, **40**, 570–583.

(*Received 15th February*, 1951.)

THE TIMING OF RESPONSES TO SPATIAL PERCEPTION QUESTIONS

BY

V. R. CANE and V. HORN

(*From the Medical Research Council Applied Psychology Research Unit, Cambridge*)

EXAMINATION of the time of response to single problems in spatial perception suggests that the approach of subjects to multiple-choice questions may differ from their approach to free-choice questions. There is some evidence for thinking that the slower subjects score more highly, and that practice in a given type of problem affects speed rather than correctness of response. These results may provide a new interpretation of some effects previously observed in repeated retesting of the same group.

The design of the experiment discussed was a "balanced incomplete block," in which 120 subjects were required to answer 108 out of the total of 540 problems.

I

AIMS OF EXPERIMENT

This research was originally undertaken as a basis for preparing a number of parallel tests of spatial perception. We wish to have such a group of tests to enable us to investigate, for instance, the effects of practice and of certain sorts of teaching. These enquiries would require the testing of the same subjects on more than one occasion, and there are obvious objections to using the same test repeatedly.

The Shapes Analysis Test, consisting of six types of questions each requiring some sort of mental manipulation of shapes, was prepared. There were 540 questions in all, 90 of each type, and our aim was to sort them out into a number of equivalent sets. We decided that two questions could be regarded as equivalent if they were of the same type, if the same number of people in a given population answered each correctly, and if the average time spent on each was about the same. This last point has not, we think, been considered before but is clearly of some importance where tests given with a time limit are concerned. It was necessary, therefore, to devise some method of recording not only the answer of each subject to each question, but also the time he spent on it. The apparatus constructed for this is described in detail later.

Since the time records were necessary in any case, we decided that we might in the same experiment make some preliminary enquiries into the relation between the speed and score of subjects, and between the difficulty of a question and the time spent in doing it. This did not in fact make the experiment more complicated, as will be seen from the discussion on the experimental design. We also gave the subjects an intelligence test so as to be able to compare their score on this with their score on the Shapes Analysis Test.

II

METHOD OF EXPERIMENT

1. *Subjects*

We chose as our subjects 13- and 14-year old children at a secondary modern school. Since it is known that the ability to deal with this sort of visual material depends on the age and sex of the subject, it seemed reasonable to take equal numbers of each sex and of two age groups from a population that was in other respects fairly homogeneous. We therefore took four batches each of 30 children, 13-year-old boys, 13-year-old girls, 14-year-old boys and 14-year-old girls, avoiding as far as possible children from the "C stream," who were the worst by the school's standards.

2. *Questions*

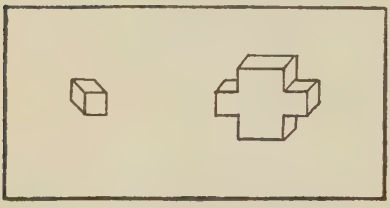
Of the six types of questions (called A, B, C, D, E and F), A, B and F were three-dimensional though represented in a plane, and C, D and E were two-dimensional. A and C required the subject to supply his own answer whereas D, E and F offered a choice of five

FIGURE 1

EXAMPLE QUESTIONS

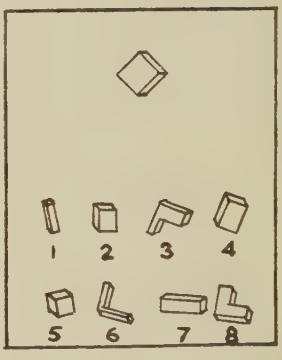
THE SHAPES MAY BE TURNED IN ANY WAY.

A 3-dimensional



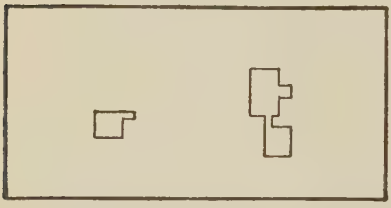
How many units like the small figure are needed to make the large figure?

B 3-dimensional



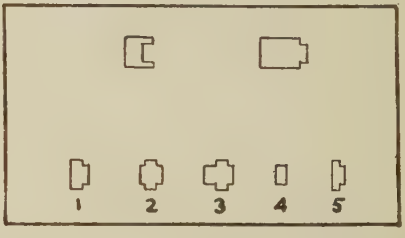
Which two of the eight lower figures are needed to make the figure above them?

C 2-dimensional



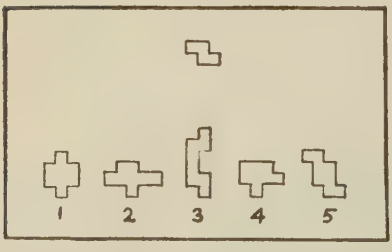
How many units like the small figure are needed to make the large figure?

D 2-dimensional



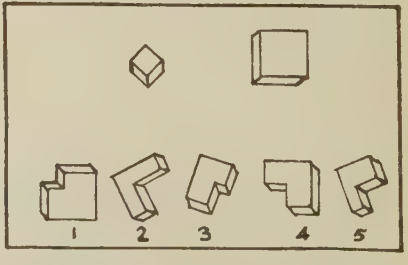
Which one of the lower figures will remain if the top left figure is taken from the top right figure?

E 2-dimensional



Which one of the five lower figures cannot be made from a number of units like the figure above them?

F 3-dimensional



Which one of the lower figures will remain if the top left figure is taken from the top right figure?

answers. B offered a choice of 28 alternative answers, and therefore it was considered to be more like A and C than the multiple-choice ones, D, E and F. The questions of each type were divided into five sets of eighteen each, balanced in difficulty as far as we could guess. The sets were called A1, A2, A3, A4, and A5; B1, etc. to F5. Each subject did six of these sets, two from each of three types, making 108 questions in all. The subject had to do examples of the relevant type of question before embarking on the test proper. These examples (some of which are shown in Figure 1) are simple illustrations of the principles concerned. The test questions themselves are in general much more difficult, as we aimed to produce a test of a very high level.

3. *Timing and Answer Recorders*

The Timing Recorder was made in the laboratory workshop. It is an ink-writing polygraph with seven pens, each controlled by a solenoid which, when energized, causes the ink line on the paper to move sideways about an eighth of an inch. One pen is used as a time marker and plots time intervals every five seconds. Each of the remaining six pens is connected to one of the Answer Recorders, thus allowing six subjects to be tested simultaneously.

The answer sheet completed by a subject is in the form of a roll of paper with small rectangular spaces ($\frac{3}{4}$ in. $\times$ $1\frac{1}{2}$ in. and $\frac{1}{2}$ in. apart) printed along the length of the paper. Each space is numbered serially to correspond with the question numbers, and the subject is asked to write his answer in the appropriate space.

The roll of paper is inserted in an Answer Recorder which has an aperture through which the subject can see only one numbered space at a time. When the subject has answered a question, he at once turns the knob at the side of the machine and so brings the next blank answer space into view.

A star wheel fixed to the same shaft as the knob operates a contact so that all the time the knob is being turned the circuit to the solenoid, controlling one of the "Answer pens," is being alternately made and broken, causing the pen to move to and fro across the paper of the time recorder. Thus the turning of the knob on any Answer Recorder will be registered by the movements of the appropriate pen, and at all other times the pen will be recording a straight line. The length of straight line between each group of to-and-fro marks is therefore a measure of the time spent on each question.

4. *Testing Technique*

The 120 schoolchildren were tested in groups of six. General instructions were first given to each group of subjects on the use of the Answer Recorders. No mention was made of the Time Recorder. Subjects were told that an answer was to be given to every question, but if, after considering a question for some time they were still uncertain about the answer, they should guess rather than omit it. It was emphasized that there was no time limit and it was clear that the children did not realize that they were being timed on individual questions.

Each child was given a book of questions and the instructions for that particular series—the first 36 questions of the 108 he was to answer (in accordance with the plan of the experiment). In addition to understanding the first two examples for which the answers were supplied, subjects had to answer correctly two other examples before starting the test proper. Individual assistance was given when necessary to ensure that every child understood the principle of the type of question he was to answer. The period of completing the example questions was useful not only in providing practice with the questions, but also for some experience in the manipulation of the Answer Recorder.

After answering the 36 questions of the first two sets, the children were given a rest and drank their mid-morning milk. They then completed the second lot of questions after which they stopped for another break and school dinner, and in the afternoon answered the third book of 36 questions. Each child spent one complete school day taking part in the experiment.

III

The Experimental Design and the Analysis involved in it

Since there are six types of questions and 90 questions of each type, and we wished to compare questions of the same type in respect of difficulty and time spent on them, it was clear that a somewhat complex experiment was necessary. It is not enough to

give the 90 questions several times over in the same order as some positions may be more favourable than others—e.g. late in the series if there is a learning effect, or early in the series if there is a fatigue effect. Moreover there may be transfer between individual questions. Ideally we should like to give the questions of one type in all possible orders, or at least so that each occurs equally often in each position; but even if we could do this, we should still have no means of comparing the difficulty of different types of questions, which is interesting if not essential to our immediate aim, nor could we make allowance for the transfer effect which may occur between types of questions, a matter which would need attention at some stage in the construction of the parallel tests. It was therefore decided that in this experiment each subject should do a series containing more than one type of question, that each question should be done an equal number of times and with an equal chance of being in any particular position in the series, and that each pair of types should be done equally often together in a series.

It was thought that a convenient number of questions for each subject to answer was 108, 36 from each of three types. We arranged thirty sets of 18 questions, as explained before, and each subject did six of these, two from each type, e.g. A1 A5, B2 B3, C4 C2. (See Experimental Design in Appendix.) To begin with, these sets, rather than individual questions, were treated as the units. The subject did two sets from each type so that practice effect could be observed directly, that is a comparison of the performances of the same subjects on the same type of questions could be made; they were done in immediate succession so as to cause as little interruption in the experiment as possible. The questions within a set were randomised by means of a random number table each time they were used. In this way, each question had an equal chance of appearing in each position in the set, and the design of the experiment ensured that each set appeared equally often in each position in the series. Also questions of type A, for instance, were preceded equally often by each other type, and were done equally often in each position by each batch of subjects. It was thus possible to assume that the total scores on different questions were comparable in respect of learning or fatigue effect, effect of transfer from one type to another, and effect of age and sex of subjects; that is, two questions of equal difficulty would not appear to be different merely because one was more often done late in a series, or by boys, or after some other type of question that was particularly favourable. It was still possible, however, that the difference between two questions in different sets would be affected by their being done by individuals of very unequal ability and an adjustment was made for this (see Appendix). Apart from this adjustment between sets of questions of the same type, all the estimates of difficulty of questions whether individually, or by types, were made from the total number of right answers to the question. The ability of the subject was estimated by his total score, with some correction made for the difficulty of the questions he had to do. The adjustments for difficulty of questions and ability of subjects were not, of course, independent, but were suitable as first approximations.

A slightly more complicated method was necessary in dealing with the time spent on questions. In the first place it was clear that the time scores were by no means normally distributed; it was therefore decided to work with the logarithm of the time. (When the histogram of this was plotted, it was found to be symmetrical and reasonably like normal.) The second difficulty was that no experimental design could arrange for balance in respect of right or wrong answers, which could appear from any subject, to any question, in any position in the series. Therefore before making a comparison between time taken by a subject and his score, or the time spent on a question and its difficulty, we wished to discover whether there was any difference

in the time taken to give a right answer and to give a wrong answer. If for instance it were quicker in general to get a test question wrong, perhaps because less attention was paid to studying it, then the subject with high scores would be slower because he had more right answers, and the difficult question would be done more quickly because it had more wrong answers; and yet it would be misleading to say that the good subject was an intrinsically slow worker or that the hard questions took less time. Thus it was necessary to work out whether there was any consistent effect of this sort before looking for other effects concerning the relation of score and time. We therefore considered separately each set of eighteen questions and each of the four groups of children doing them and, after making an adjustment for the difference in performance of the subjects and difference in average time of the questions, compared the total time spent on right answers with the total time spent on wrong ones. There was very rarely any significant difference between these two times, and as the effect was not consistent, the right answers sometimes taking less and sometimes more time, and as also the average effect, for each type of question, was insignificant, we felt that we were justified in neglecting even the significant cases.

The rest of the analysis of the time measures could then be made in the same way as that of the scores.

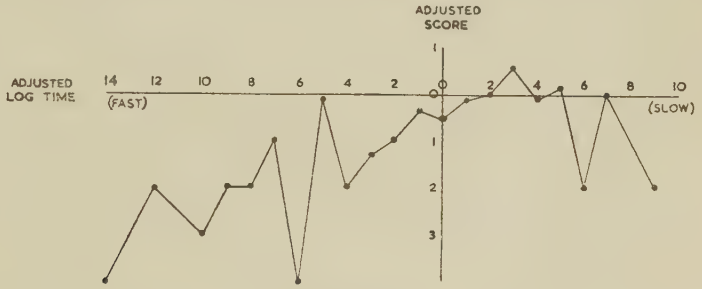


FIGURE 2

REGRESSION OF SCORE ON TIME (AVERAGE SCORE OF SUBJECTS TAKING THE SAME TIME)

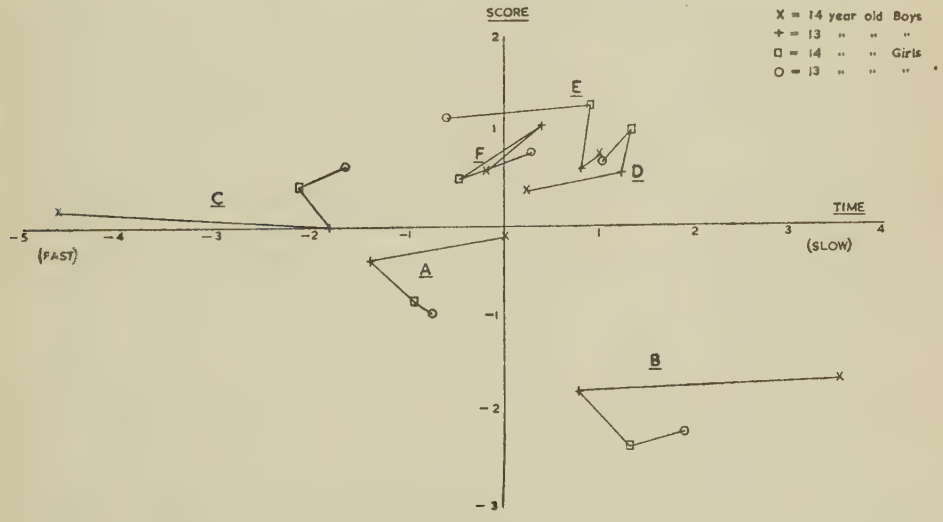


FIGURE 3

RELATION OF LOG TIME TO SCORE FOR QUESTION TYPES

IV

Results

At the end of the analysis we had the following estimates:

(i) The score and speed for each subject,
(ii) the difficulty (within its type), and the mean time taken, for each question,
(iii) the relative difficulty of, and time taken by, the different types of questions,
(iv) the total effect of practice on score and on time, not distinguishing between types of questions,
(v) the time and score for batches of subjects of different age and sex.

(i) and (ii) were plotted on scatter diagrams, which are not given here. The regression line of Figure 2 summarises the relevant information about the relationship of score and speed of subjects, and a set of two by two tables (see Appendix, Table I) shows how far questions which are above or below average difficulty take more or less than average time.

The scatter diagram for (iii) is given above (Figure 3). For each type of question, and each batch of subjects, the average number or correct answers is plotted against the average log-time taken. Thus there are four points corresponding to each type, and these are linked by straight lines so that the difference between types can be readily seen.

On the results of (iv) and (v), two analyses of variance were performed, one for score and one for time (see Appendix, Tables II and III).

It was found that the position of a set of questions did not affect its score, i.e. there was no fatigue or practice effect. There was, however, not only a significant difference between the scores of boys and girls, and between 13- and 14-year olds, as was expected, but also a significant age and sex interaction. In fact from the following table of total scores,

	Boys	*Girls*
13-year old	1040	786
14-year old	1095	979

it can be seen how much more girls improve than boys between 13 and 14.

The time spent on a set of 18 questions was significantly affected by its position, in fact the time decreased by about one-tenth on the average between one position and the next. The drop was particularly noticeable between positions 1 and 2, 3 and 4, 5 and 6, when subjects were doing two sets of the same type, but there was also a drop from positions 1 to 3, and 2 to 4, so that there was a general, as well as a particular, practice effect on *speed*. There was again a difference between boys and girls (boys being slower), but no significant age difference.

It should be pointed out that the test questions were very difficult for the children; the scores were generally less than half the possible maximum and for type B questions many of the scores were zero. This was not surprising, since the test has presented considerable difficulty even to University students.

V

Discussion

The most interesting results of this experiment are those involving the time measurements, for, so far as we know, very few systematic enquiries have been made into the relation of time taken and correctness of answer, and those that have been undertaken have generally been for a rather different purpose. For instance, Peak

and Boring (1926), McFarland (1930) and Slater (1938) have all shown that subjects tend to be consistently fast or slow in their responses, whereas we confined ourselves to considering the average rate of response of a subject. Macleod (1929) concludes that the level of difficulty of a task affects the correlation between speed and accuracy; we have only concerned ourselves with the relation at a constant level of difficulty, and that rather a high one. If what we have found with this particular series of tests should prove to be true with all sorts of tests, many of the effects which till now have been expressed in terms of score—as, for instance, that on timed tests the better subjects, in certain circumstances, improve more with practice than the poorer ones—may more satisfactorily be expressed in terms of time, and it may even be possible to connect time spent with some set of decisions that have to be made between alternatives. For example, when we investigated whether it took longer to get an answer right than to get it wrong, for similar subjects on questions of similar difficulty, we found that, although the differences in time were insignificant for each type considered separately, on the average it did take longer in the case of multiple-choice questions and less time in the case of questions to which the subject had to supply his own answer. (In fact, when the results for questions of the latter type were pooled, correct answers took *significantly* less time than incorrect ones. The effect, however, was small, the ratio of the two times being about 1·06.) This would agree with the idea that in the multiple-choice case the right answer would more likely be found if all the alternative answers were considered, whereas in the free-choice case the right answer would be found when the subject was offering himself a set of reasonable possibilities only, successfully limiting his field of enquiry to cut out unprofitable alternatives. It should be possible to devise experiments of a kind other than the paper-and-pencil test to explore this.

Two results which are most interesting, especially when considered together, are (i) that practice considerably diminished the time spent on such a test (in fact, by nearly 40 per cent. between the first and sixth sets of eighteen questions), but had little if any effect on the score, and (ii) that the better subjects tended to have a lower average speed. If this applies in general (at least with tests difficult for the group concerned), it means that when a test with a time limit is first given, the poorer subjects will cover more of the test in the given time and so approach nearer to their maximum score than the better subjects do. If, now, the test is given again to the same subjects they will all tend to work faster, but the increase in the number of questions answered will be of greater benefit to the good subjects, who will get more of them right, so that the better subjects will improve their scores more than do the worse ones. This may possibly be the explanation of the results found in repeated retesting with the same timed intelligence test (Cane and Heim, 1950) for an examination of the graphs given there shows that

(i) the increase in the number attempted is at least 55 per cent. over ten testings, whereas the increase in ratio of correct to attempted is at most 16 per cent., i.e. the most important effect of practice is improvement in speed,

(ii) better subjects work more slowly than poor ones on a difficult test.

The existence of a time limit in such a test will thus tend to minimise the differences between subjects, but this can be partly offset by giving the subjects preliminary practice on a similar test. If, in addition, it should be generally true that it takes longer to get a multiple-choice question right than wrong, then the preferential treatment which a test with a time limit extends to the weaker subject would probably be even greater in a test composed entirely of multiple-choice questions.

The remark in the previous paragraph, that better subjects tend to work more

slowly than poorer ones, needs a little modification. As will be seen from Figure 2 the relation is not linear, the very slowest subjects (at the right hand side of the graph) showing a decline in score. Similar effects have been observed when intelligence tests have been given without a time limit (Cane and Heim, *in preparation*).

Among questions of the same type we found that the hardest questions tend to take longer, though the relationship is not very strong even for questions where the subject has to invent an answer and is less so when he merely has to choose one (see Appendix). When the time and score for different types of questions are compared (see Figure 3) it can be seen that the multiple choice questions again fall into a separate category—they have scores above the average, take about average time, and are very like each other with respect both to score and to time. The other three types of questions, C, A and B lie in very different parts of the plane of the graph, and roughly lie on a line which suggests that the most difficult types have the most time spent on them. This again suggests that in material of this type at least, the discriminatory power of multiple-choice questions is much less than that of free-choice ones.

The boys scored higher than the girls on each type of question, the greatest differences occurring with the three-dimensional types, A, B and F, but the difference between the older boys and girls was much less than between the younger boys and girls, so that girls probably improve between 13 and 14 to a greater degree than boys. It has been found before (Harris, 1947–49) that the intelligence test scores of girls increase more from 13 to 14 than in any other year between 11 and 16.

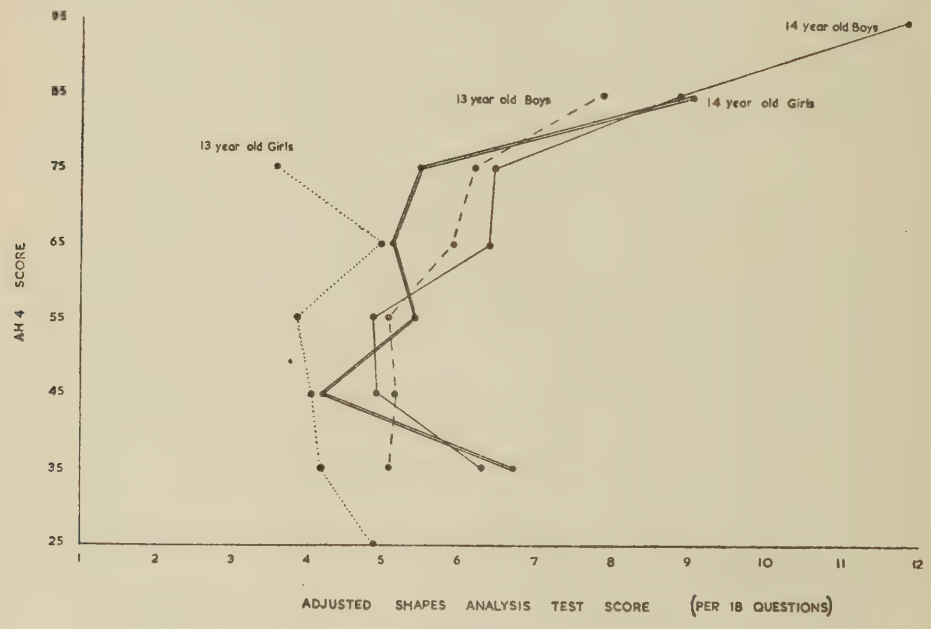


FIGURE 4

Relation of Shapes Analysis and Intelligence Test Scores

There was a slight association between adjusted Shapes Analysis Test score and score on the intelligence test AH 4, the average correlation for the four groups of children being ·29. The distribution of AH 4 scores for these children was similar to that for an unselected industrial population except that the worst 10 per cent. of subjects had been omitted by the method of selection; for this reason we show also

a regression line. It is clear from Figure 4 that the relationship is by no means linear, but that subjects scoring highest and lowest on AH 4 do rather better on the Shapes Analysis Test than would be expected from the trend in the middle of the group. The fact that the Shapes Analysis Test was so much too difficult for the weaker subjects, as well as some selection in the subject range, has undoubtedly reduced the correlation; when many of the scores are zero (as occurred especially with type B questions) there is no discrimination at all between subjects of medium and low ability, and their total Shapes Analysis Test scores will tend to be equal. This may account for the almost complete lack of relationship between the two scores in the middle of the group, but it cannot explain the relatively good performance of the six worst AH 4 subjects, nor can it explain why the younger girls score least on the Shapes Analysis Test at each level of score on AH 4. It is difficult indeed to give any plausible reason for this; but as the correlations between the two scores are greater for the older children than for the younger it definitely does not support the theory that there is an increasing differentiation of abilities with age. Altogether it appears that the relationship between our test and an accepted intelligence test is very slight; how far they differ because the two tests call upon different "factors" in the usual sense, or because one is given in limited time and the other is not, or because one is very much more difficult than the other, cannot as yet be decided.

VI

Conclusions

It is clear from the results of this experiment that there are time effects which would play an important part in determining the score of a subject in a timed test. Since the relation between performance on this test and on a timed intelligence test is very slight, it cannot be asserted that similar time effects would be observed with intelligence test material, but the nature of the results make it likely that it would be so, and it should probably be investigated.

In particular, it appears possible that the number of alternative responses provided for a question may influence the time spent on it, and that when there is complete freedom of choice the behaviour of the subject may be of an essentially different kind from what it is in a situation involving limited choice. This could probably best be investigated by observation of individual subjects, and we hope to do some experiments with various sorts of material with and without a time limit, and with and without instructions to guess an answer when in doubt. We shall moreover reconsider all these conclusions in the light of the results we get when we use the parallel tests prepared from these questions.

We wish to express our thanks to Dr. A. W. Heim, who has given us every possible support in this experiment, to Capt. A. M. N. Aston, M.C., and his staff and pupils of the Eldon Road Secondary Modern School, Edmonton, who so willingly co-operated in the actual testing, and to Mr. F. J. Anscombe of the Statistical Laboratory, Cambridge, who both produced the experimental design and gave us most valuable advice at every stage of the analysis of our results.

VII

References

1. Cane, V. R., and Heim, A. W. (1950). The effects of repeated retesting: III. Further experiments and general conclusions. *Quart. J. exp. Psychol.*, **2**, 182–197.
2. Cane, V. R., and Heim, A. W. The effect of time limit on test score. In preparation.
3. Harris, N. (1947–50). Reports on testing and retesting grammar school girls at yearly intervals. Unpublished.

4. McFarland, R. (1930). An experimental study of the relationship between speed and mental ability. *J. gen. Psychol.*, **3**, 67–97.
5. McLeod, L. S. (1929). The interrelations of speed, accuracy and difficulty. *J. exp. Psychol.*, **12**, 431—442.
6. Peak, H., and Boring, E. G. (1926). The factor of speed in intelligence. *J. exp. Psychol.*, **9**, 71–94.
7. Slater, P. (1938). Speed of work in intelligence tests. *Brit. J. Psychol.*, **29**, 55–68.

(*Received 1st March*, 1951.)

STATISTICAL APPENDIX

Design of the Experiment

	13-year-old Boys			*14-year-old Boys*			*13-year-old Girls*			*14-year-old Girls*		
1	E45	C52	D43	F35	B51	C25	F32	A25	C23	F31	D15	B41
2	E12	A13	C32	A41	C34	B25	D21	A42	C12	F42	C43	A24
3	F12	B52	E54	C52	B12	A53	D43	B42	C34	A53	F15	B24
4	A34	C13	E21	C41	E45	D31	F21	D25	E34	F14	C54	E52
5	D34	C24	F21	F13	E51	B31	D32	E42	F12	D14	C32	A35
6	A12	D13	B32	C24	F12	B42	A32	B31	F23	D53	B32	F13
7	D51	E52	B54	E24	A51	F53	A15	E25	D45	C31	A15	F35
8	F23	D52	C15	D41	B45	E25	B15	F31	A34	E31	F54	C41
9	C23	E41	A54	A52	B23	D25	A21	C25	D12	B42	D43	C35
10	C12	A41	B15	E41	B34	F42	C54	D42	B51	C25	E43	F41
11	C51	F35	D32	E13	C45	A42	C43	B53	E45	E53	D54	F24
12	A45	F41	D21	B35	C12	F25	B54	D31	F34	A31	D32	E24
13	C34	D35	E32	F41	C23	D53	F15	A53	B23	B53	A54	F52
14	E23	F52	A43	E35	D45	C42	A43	C31	F45	B14	C15	E13
15	A51	B13	C54	C35	A12	E14	E21	C14	B12	E14	B15	C24
16	E51	F24	B21	A35	F34	E53	C15	D53	A51	A25	F21	C13
17	F45	C35	B43	B24	A34	C53	B21	E31	C51	D42	C21	B35
18	D23	A52	F43	D13	F45	A14	F43	E53	C45	F25	E15	D13
19	D12	E24	C21	C13	D23	F31	E15	C53	F51	E42	A43	D24
20	D45	B24	A21	E52	D34	B14	C32	F53	A12	E25	A32	B13
21	A23	E13	F54	D52	C51	E31	E43	D14	A23	A42	D21	C52
22	F34	A35	E43	A24	D12	F14	C21	F25	E51	C14	E54	B52
23	E34	B35	D15	D35	F51	C31	F54	B25	D51	F53	B54	A52
24	C45	B41	F32	F52	E23	A31	D15	F14	B45	D25	F32	E41
25	B23	C41	A32	D24	A23	B53	E54	B14	A45	C53	B43	D41
26	F51	D24	A15	F24	A45	D42	D54	A31	E23	B31	E21	A13
27	B34	D41	E15	B41	F23	E42	A54	E14	B34	B25	F43	D52
28	B45	A24	D54	B13	E34	D14	B43	A14	E12	D31	E32	A41
29	B51	E35	F15	A13	E12	C14	B32	C42	D34	A14	B21	E35
30	B12	F13	C43	B52	D51	A25	E32	F42	D23	C42	A21	D35

I

Analysis of Scores

For each of the four batches of children, we estimated the main effect of, for example, type A questions by

$$a = \mathrm{A} - \frac{\alpha}{3}$$

where A = sum of scores for type A questions,

α = sum of scores of subjects doing type A questions;

then $\Sigma a = a + b + c + d + e + f = 0$.

We estimated also a correction for set effect, for boys and girls separately, by, for example,

$$a_1 = \frac{1}{15}\Sigma(a_1 - a_k)$$

where A1AK, or AKA1, are a pair of sets done by the same subject, summed over all such pairs including A1;

$$\text{then } \sum_i a_i = 0$$

The effect of the set A1 was then taken as $a + a_1$, and similarly for the other sets.

The score effect s_i, for the i^{th} subject, was obtained for each batch separately from

$$s_i = \frac{1}{6}S_i - \frac{1}{180}G - \frac{1}{6}(\text{sum of effects of sets done by this subject})$$

where S_i is the subject's observed score and G is the total score of subjects in that batch. (s_i is thus an estimate of the subject's score on an average set of 18 questions).

The difficulty of the k^{th} question in the set A1 was estimated for boys and girls separately by

$$q_k = \frac{1}{12}Q_k - \frac{T_1}{12 \times 18} + \frac{a_1}{18}$$

where Q_k is the total number of right answers to the k^{th} question (from 12 children) and T_1 is the total number of correct answers to the 18 questions in this set.

Then $\Sigma q_k = 0$ summing over questions of the same type (actually, we worked with $12q_k$).

II

Analysis of Times

The times were read from the recording rolls by a scale which transformed them to logarithms to the base $10^{\frac{1}{2}}$; there was also an arbitrary scaling factor of about $\frac{3}{35}$, so that what we shall henceforth call a time is in fact $2 \log_{10}$ (time in seconds) $- 2{\cdot}1$.

Since the experiment could not be balanced with respect to the appearance of right and wrong answers, the times were first examined to see whether a right answer took a longer or shorter time than a wrong answer. For each set of 18 questions, which were done by six subjects from each batch, we fitted a constant γ by taking the maximum likelihood solution of

$$t_{ij} = \mu + \alpha_i + \beta_j + \gamma + \epsilon_{ij}$$

where t_{ij} is the time spent by the i^{ht} subject on the j^{th} question, and γ is omitted if the question is answered incorrectly, so that γ is a measure of the extra time taken to get a question right.

$$\text{then } \hat{\gamma} = \frac{T_r - \frac{\Sigma N_i T_i}{18} - \frac{\Sigma N_j' T_j'}{6} + \frac{NT}{6 \times 18}}{N - \frac{\Sigma N_i^2}{18} - \frac{\Sigma N_j'^2}{6} + \frac{N^2}{6 \times 18}}$$

where N = number of correct answers,
T_r = total time of correct answers,
T_i = total time for i^{th} subject,
T_j' = total time spent on j^{th} question,
N_i = total score of i^{th} subject.
N_j' = number of times j^{th} question was answered correctly.

There were 20 such $\hat{\gamma}$'s for each type of question, one from each of four batches and each of five sets, so that altogether there were 120. We tested eight of these by using

$$\text{var } \gamma = \frac{\text{residual mean square}}{N - \frac{\Sigma N_i^2}{18} - \frac{\Sigma N_j'^2}{6} + \frac{N^2}{6 \times 18}}$$

and found one significant at the 1 per cent. level (this was in fact the $\hat{\gamma}$ with the highest value), one significant at the 5 per cent. level and the rest not significant. There was, however, no consistent arrangement of the values obtained; we had expected, for instance, that sets of the same type and done by subjects of the same batch would yield similar results, but nothing of this sort was detected either by inspection or by analysis of variance. The average value of $\hat{\gamma}$ for the whole set of 120 was $-\cdot 019$, and the standard deviation calculated from this set was ·13 (which was very similar to the standard deviations calculated in the eight individual cases mentioned above). We therefore decided that the effect, if it existed, was not of sufficient importance in this case for us to need to make allowance for it. It should be noted, however, that the three types for which $\hat{\gamma}$ was on the average greater than 0 were those in which a choice of answers were offered.

The rest of the analysis of the times was the same as that for the scores; an estimate t_i of the time spent by the i^{th} subject on an average set of 18 questions corresponded to the previous s_i, and an estimate l_k of the length of time spent on the k^{th} question by an average subject corresponded to q_k. The relation of s_i to t_i, of s_i to the score on the intelligence test, AH 4, and of q_k and l_k were shown on scatter diagrams.

Estimates of the effect of position in the series of sets of questions, of sex, age, type of question, or any interaction of these were made by adding the relevant observed scores or times, since the balance of the experiment allowed this.

TABLE I

Relationship of Difficulty of Question and Time Spent on it

Type		*Difficult*	*Easy*	$\sqrt{\frac{\chi^2}{N}}$
A	Slow	69	24	·37
	Fast	33	54	
B	Slow	62	34	·12
	Fast	44	40	
C	Slow	79	19	·47
	Fast	28	54	
D	Slow	58	39	·15
	Fast	37	46	
E	Slow	52	44	·05
	Fast	41	43	
F	Slow	44	40	·04
	Fast	46	50	

TABLE II

Total Scores

	Position						
	1	2	3	4	5	6	*Total*
14-year-old Boys	188	200	181	181	175	170	1095
13 ,, ,, ,, ...	170	191	171	170	178	160	1040
14 ,, ,, Girls	175	167	145	153	165	174	979
13 ,, ,, ,, ...	129	125	139	124	140	129	786
Total ..	662	683	636	628	658	633	3900

Analysis of Variance

	s.s.	*d.f.*	*m.s.*	*Var. ratio*
Positions	566·5	5	113·3	Not sig.
Sex	5704·2	1	5704·2	58·09 } sig. at ·1% level
Age	2562·7	1	2562·7	26·10 }
Sex x Age	793·5	1	793·5	8·08 sig. at 5% level
Residual	1473·1	15	98·2	
Total ..	11100·0	23		

TABLE III

Total Time Measures

	Position 1	2	3	4	5	6	*Total*
14-year-old Boys ..	811·5	674·1	681·2	577·3	656·4	549·6	3950·1
13 ,, ,, ,, ..	833·7	685·5	692·6	568·0	682·1	579·6	4041·5
14 ,, ,, Girls ..	805·3	670·7	649·2	508·6	629·7	551·9	3815·4
13 ,, ,, ,, ..	803·3	677·1	635·8	542·2	639·7	597·0	3895·1
	3253·8	2707·4	2658·8	2196·1	2607·9	2278·1	15702·1

Analysis of Variance

	s.s.	*d.f.*	*m.s.*	*Var. ratio*
Positions	176881·62	5	35376·32	
Regression	113572	1	113572	393 } sig. at ·1% level
Deviation from regression	63310	4	15828	55 }
Sex	3292·38	1	3292·38	10·3 sig. at 1% level
Age	1219·80	1	1219·80	Not sig.
Sex x Age	5·70	1	5·70	,, ,,
Residual	4290·07	15	286·00	
Total ..	185689·58	23		

Regression slope, $b = -40·3$.

MISCELLANEA

A NOTE ON THE PAIRED COMPARISONS METHOD OF RATING

BY

NANCY HARRIS

(*Medical Research Council Applied Psychology Research Unit, Cambridge*)

In psychological work where assessment of the ability of a number of individuals is required, paired comparisons are often preferable to any other method of rating. But if this method is used, the number of subjects rated in any one group must be kept small; 20 subjects, for example, will involve the rater in 190 comparisons.

When a large group of subjects is to be assessed, it is usually possible to sub-divide it for purposes of rating. But the nature of the division is often dictated by circumstances outside the investigator's control, and it is common to find that the resulting sub-groups are of unequal size. The data from a number of unequal sub-groups may usually be analysed separately, and the results summated to give a figure for the whole group. χ^2 and $\hat{\imath}$, for example, are appropriate statistics to use in this way.

Sometimes, however, a single set of ratings on the whole group of subjects is exacted by the aims of the research. The following method of recombining the paired comparisons data from a number of sub-groups was found useful in a recent investigation where this situation arose.

Suppose that a group of 30 subjects are to be assessed on proficiency in a certain job. The available raters are such that the group can be sub-divided in one way only—that is into four sub-groups numbering, respectively, 6, 6, 9, 9. Each of these groups can be assessed by three independent raters. For each group, the number of times each subject is judged the better of a pair is counted, and preference scores are obtained in this way for every subject. Now if the raters were all perfectly self-consistent and if the three raters for each group were in perfect agreement with each other, we should obtain the following sets of total preference scores:

Subject:	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*J*
Sub-groups I and II (6 subjects in each)	15	12	9	6	3	0			
Sub-groups III and IV (9 subjects in each)	24	21	18	15	12	9	6	3	0

In fact we find that every rater is highly self-consistent but not perfectly so, and the agreement between the three raters for each sub-group is, in every case, significant but not perfect. The observed preference scores for the four sub-groups are as follows:

Subject:	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*J*
Sub-group I	15	10	8	8	4	0			
,, ,, II	13	12	11	4	3	2			
,, ,, III	24	21	17	14	12	10	7	3	0
,, ,, IV	22	20	20	16	12	8	5	4	1

It is possible to match these observed preference scores with the theoretically perfect ones. We can take the smaller sub-groups (I and II) as defining the number of grades of differentiation and regard the perfect preference scores as the boundaries of five grades as follows:

Grade 1	12–15
Grade 2	9–12
Grade 3	6–9
Grade 4	3–6
Grade 5	0–3

We can fit the sub-groups I and II into these grades directly, adopting the convention of putting scores which coincide with a boundary figure into the grade nearer the middle. For the larger sub-groups, III and IV, a conversion must be made. The maximum theoretical preference score for these groups is 24, and by dividing this by 5 the size of the grades and the boundary figures are determined. They would be:

Grade 1	19·2–24·0
,, 2	14·4–19·2
,, 3	9·6–14·4
,, 4	4·8– 9·6
,, 5	0·0– 4·8

The following classification of the observed data can then be made:

Sub-group:	*I*	*II*	*III*	*IV*
Grade 1 ..	A	A	A, B	A, B, C
,, 2 ..	B	B, C	C	D
,, 3 ..	C, D	—	D, E, F	E
,, 4 ..	E	D, E	G	F, G
,, 5 ..	F	F	H, J	H, J

As the theoretically perfect preference scores always follow a regular series, this procedure can be used to classify the whole group into any number of grades, up to the point when the grading approximates to a ranking.

The method described enables data from sub-groups of different sizes to be combined, and it preserves differences of distribution within the groups. The main criticism is that it assumes every sub-group to be of similar overall standard. When all the sub-groups are drawn from the same source it may be possible to arrange that this is true; failing that, it may be possible to select members from each sub-group and by having them assessed in a further mixed group to check up on the relative standards of the sub-groups. But when the sub-groups come from different sources there is no way of estimating how far they are of a common standard. This criticism would, of course, apply to data obtained by any method of rating and combined in any way. By using this method of treating paired comparisons one can at least be certain that the average level of classification grade 1 is higher than that of grade 2, and so on, even if the extent of variation *within* any one grade is unknown.

BOOK REVIEW

The Perception of the Visual World. By J. J. Gibson. Boston: Houghton Mifflin. 1950. Pp. xii + 235.

In his introduction to this book, Professor Carmichael says that "the study of perceptual problems is central in psychological theory." The reviewer is in complete agreement with that statement; and there can be no doubt of the value of Professor Gibson's contribution to this study. One may not be able to agree completely with his argument—and indeed its exposition is not always as straightforward as might be wished. But it is certain that everyone who is interested in this central problem of psychological theory should study that argument.

Professor Gibson presents a theory of perception which endeavours to resolve the old empiricist-nativist controversy, and to reconcile the differences between the Gestalt psychologists and those who hold that the primary sensory data must be developed and interpreted in the light of experience before meaningful percepts are obtained. He is also concerned to controvert recent theorists who have maintained the view that perception is determined to a considerable extent by individual personality factors, and that the perceived world is an individual construct. This, he says, cannot be so, since the visual world differs comparatively little between different observers, and agrees remarkably well with what is actually known as to the nature of the environment. "The properties of the literal visual world, in so far as conditions are optimal, approach an exact correspondence with variables of physical stimulation." But this "visual world" must be distinguished from the "visual field"—what can be observed in introspective analysis at any one moment by the single stationary eye, and what has sometimes been thought of as the primary sensory impression. The visual field consists of flat patches of colour, bounded by the periphery of vision; it corresponds fairly closely with the pattern of retinal excitation, though the latter is a correlate, not a picture, of the external field. But the visual world is continuous, unbounded, possessing depth and solidity; it is built up from the primary memory images of successive and overlapping retinal impressions, obtained by changes of fixation which are so frequent and so randomly directed that no awareness of succession remains.

It is clear that evidence of the truth of this theory must be based upon experimental observation of the manner in which varying impressions of the visual field are built up from varying visual fields. This Gibson proceeds to do by considering the phenomena of form and of the arrangement of objects in space. He states that the Gestalt laws of organization of form are on the whole misleading because they were obtained from the study of isolated and often ambiguous two-dimensional shapes. Re-organization of the field in accordance with the Gestalt laws does not occur with ordinary solid objects distributed in space; their shapes are clear, unambiguous and rigid. The figure-ground experience is also quite unambiguous, being given by the edges of these objects perceived from simultaneous sudden changes in gradients of colour, brightness, texture and distance.

The apparent stability of the visual world and of objects in space is explained as follows: When the eye or the head move, the order of arrangement of excitations is maintained temporarily and spatially; and the movement of this orderly pattern across the retina is compensated by kinaesthetic sensations of movement of the eyes and head. Thus the world does not appear to move or change. In the perception of real movement, there is again an orderly succession of changing retinal impressions, but the kinaesthetic sensations do not compensate exactly.

The perception of distance Gibson attributes to the occurrence of a regular serial order of surfaces along the ground, graded from near to far in texture, size, slant of shape (perspective), hue, clarity, disparity and deformation with movement (parallax). Gibson's work on the varying deformation of the perceived landscape with aircraft movement has shown how valuable this phenomenon can be in the estimation of distance. When no perception of receding surfaces is possible, localization is difficult if not impossible. (This is demonstrated by a neat experiment described later in the book in which it is shown that judgment of size depends on the apparent distance as given by the part of a receding surface with which the test object is in contact.) But in contradiction to Gibson's theory, there is much evidence to show that relative localization at least is possible with disparity functioning in the absence of other perceptual data. The impression of distance is based, not on the accurate gradation of disparity from near to far, but on the correspondence of the retinal images of points equidistant with the fixation point, and non-correspondence of the images of other points.

The distinction between the visual field and the visual world seems at first to come in appositely when we consider the variations of colour, shape and size constancy which have been observed to occur when there is a change from the "object" to the "stimulus" attitude. The latter of course gives the observer the analytical impression of the visual field. With the former, he observes the visual world, and his shape perception is integrated with his perceptions of the orderly distortion of outline and surface texture with tilt of the surface. But Stavrianos has shown that shape constancy is maintained when judgment of tilt is quite inaccurate. And recently Langdon, at Oxford, has gone further and shown that it is independent of texture or brightness differences. It must be noted that Gibson is somewhat too much inclined to present *prima facie* pronouncements as to the importance of factors such as surface texture without adequate experimental evidence of their actual functions.

Size constancy Gibson attributes to the fact that the size of any given object is given by the scale of the background, in the total visual world, at the point to which the object is attached. Thus the results of Joynson's experiments, showing that constancy may be impaired by observing the visual angle covered by the object, would presumably be due to the fact that the attachment is to a narrow segment only of the momentary visual field. Some objection must be made to Gibson's contention that the size constancy of objects holds as long as they can be seen at all. Any one who has sat on the side of a mountain and watched beetle-like cars and ant-like men crawling along a road in the valley will agree how genuinely minute they appear. Or is this attributable to the lack of gradually receding surfaces between the observer and the objects?

Finally, it is clear that the visual world possesses another characteristic as basic and fundamental as those of form and spatial arrangement—that of meaning. It seems probable that all perception, from birth onwards, has meaning of a kind, though the meanings attributed by the infant to his percepts are very different from those utilized by the adult, which are schematically determined from what he learns about the environment. Gibson has less to say on the subject of meaning. But there seems no sort of reason why he should postulate that the existence of such schematically determined meanings is incompatible with accurate perception of what he calls the "literal world," the properties of which "approach an exact correspondence with variables of physical stimulation." From his earliest years the child learns that in order to come to terms with his physical and social environment, he must perceive at least certain aspects of the world with sufficient accuracy to constitute a basis for understanding their effective meaning. The normal type of percept is then a construct in which are integrated impressions (colour, shape, etc.) correlated with variations in the physical variables, and meanings for the most part laid down schematically through active "trial and check" and social training. Individual meanings are found to occur when the physical conditions are inadequate to provide clear and unambiguous impressions. Then, as Bruner has pointed out, the individual is bound to make hypotheses which will be based on individual knowledge and interest in so far as the habitual schemata which he possesses in common with others are inadequate to provide meaning. Any acceptable theory of perception must account for the discrepancies as well as for the agreement of perception between different individuals. M. D. V.

OCCUPATIONAL PSYCHOLOGY

Contents of July, 1951, issue

Selection, Training and Status of Supervisors. II. Training
by F. I. de la P. Garforth, H. F. Lock and D. M. Sidney

A Study of Some Attitudes to Factory Work
by J. Walker and R. Marriott

Relation Between Skill and Adjustment of a Group of West Indian Negro Workers in England
by Anthony H. Richmond

Current Research Needs in Vocational Guidance
by Dennis McMahon

Predicting the Occupational Adequacy of Certified Mental Defectives
by N. O'Connor and J. Tizard

Predicting Aptitude for Dentistry
by B. G. R. Moore and E. A. Peel

Book Reviews

Annual Subscription £1

NATIONAL INSTITUTE OF INDUSTRIAL PSYCHOLOGY
14, WELBECK STREET, LONDON, W.1.

THE BRITISH JOURNAL OF PSYCHOLOGY

(GENERAL SECTION)

Edited by D. W. Harding

Vol. XLII, Part 3 **August 1951** **12s. 6d.**

FOULDS, G. A. Temperamental differences in maze performance.

HERON, ALASTAIR. A study of temperament in adolescence.

HUMPHREY, GEORGE. There is no problem of meaning.

LITWINSKI, LEON. Toward the reinstatement of the concept of the self.

SMITH, B. BABINGTON. On some difficulties encountered in the use of factorial designs and analysis of variance with psychological experiments.

BERLYNE, D. E. Attention to change.

BROWN, R. C. The identification of morse signals when heard against a continuous tone.

GUNTER, RALPH. Visual size constancy in the cat.

WARBURTON, F. W. Critical Notice.

Publications recently received.

The subscription price per volume, payable in advance, is 40s. net (*post free*).

Subscriptions may be sent to any bookseller or to the

CAMBRIDGE UNIVERSITY PRESS

Bentley House, Euston Road, London, N.W.1.

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY

¶ *The Quarterly Journal of Experimental Psychology* is published for the Experimental Psychology Group by W. Heffer & Sons Ltd., Petty Cury, Cambridge, England, to whom all business communications should be addressed.

¶ Papers submitted for publication and all other communications on Editorial matters should be addressed to D. Russell Davis, M.D., Psychological Laboratory, Downing St., Cambridge, England.

¶ Contributors are requested to keep the number of diagrams for publication to the minimum necessary, and all diagrams must be submitted in a form suited to direct reproduction without redrawing, i.e. they should be drawn in black ink on white card or heavy paper about 8″ × 5″.

¶ Contributors receive 20 copies of their paper free. If they require further copies they should communicate with the Publishers at the time of returning final proofs to the Editor.

¶ The subscription for *The Quarterly Journal of Experimental Psychology* is £1 10*s*. 0*d*. (post free) per volume of 4 numbers. Numbers appear quarterly. Single copies may be obtained at 8*s*. 6*s*. (postage 2*d*.). Subscriptions and orders should be placed with W. Heffer & Sons Ltd., or any bookseller.

PRINTED BY W. HEFFER & SONS LTD., CAMBRIDGE, ENGLAND